PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.
JAN 27 1923

Journal of Educational Research

OFFICIAL ORGAN OF THE
EDUCATIONAL RESEARCH ASSOCIATION

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
FRANK N. FREEMAN

VOL. VII JANUARY, 1923 No. 1

CONTENTS

	Page
PUBLICITY WORK FOR BETTER SUPPORT OF RURAL SCHOOLS — Carter Alexander	1
THE STANDARDIZATION OF TESTS AND SCALES — Peter Sandiford	14
INTELLIGENCE OF TEACHERS IN TRAINING — EFFECTS WITH INTELLIGENCE TESTS — Samuel Renshaw	28
STUDY OF COMPREHENSION IN READING BY MEANS OF A PRACTICE EXPERIMENT — Arthur I. Gates	37
INDIVIDUAL LABORATORY VERSUS DEMONSTRATION METHOD IN TEACHING PHYSICS — E. W. Kiebler and C. Woody	50
EDITORIALS	59
REVIEWS AND ABSTRACTS	63
NEWS ITEMS AND COMMUNICATIONS	75
EDUCATIONAL RESEARCH ASSOCIATION	87

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1923, by
Public School Publishing Co.

Official Organ of the Educational Research Association

Editor-in-Chief, B. R. Buckingham
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ohio State University, Columbus, Ohio.*

W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

FRANK N. FREEMAN, *Ex-Officio, Secretary Educational Research Association.*

GEORGE D. STRAYER, *Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California.*

GUY M. WHIPPLE, *University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Nebr.*

M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota, Minneapolis.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin, Madison.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean of University Administration, University of Kansas, Lawrence, Kans.*

ROBT. H. LANE, *Assistant Superintendent of Schools Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.*

LEON O. SMITH, *Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.*

W. W. THEISEN, *Director, Division of Reference and Research, Public Schools, Cleveland, Ohio.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Mich.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Co. The subscription price is $4.00; $4.10 in Canada or Mexico; and $4.20 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on editorial matters to B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume VII JANUARY, 1923 Number 1

PUBLICITY WORK FOR BETTER SUPPORT OF RURAL SCHOOLS

CARTER ALEXANDER

For years the advocates of public education have been troubled about the rural school. To them an important goal of education has been approximate equality of educational opportunity, and they have known that country children in most places have had nothing like the opportunities for schooling which are open to city children in the same sections. Whatever the biases of these advocates, they have almost invariably stipulated that the rural schools of the state must have greater revenues before the rural children can have an equal chance at education. But probably not one such advocate in a hundred has thrown a spark of light on the crucial problem involved. This problem is how to get the rural taxpayer, who has always borne and must for some time to come continue to bear the bulk of the expense for rural schools, to be willing to pay more taxes, to agree to a larger school district, or to vote for a tax reform that will produce greater revenues for schools. Until accurate knowledge of how to influence the country taxpayer favorably to these ends is secured, pronouncements on what ought to be done for rural-school support are for all practical purposes merely pious wishes.

In this situation, and perhaps because this is an age of publicity, it is natural to think of what might be accomplished by the right kind of public appeal. But although publicity work has had signal success in securing better support for schools and for numerous other social institutions in cities, very little is known

of its possibilities for helping country schools. In a few scattered places local or state publicity campaigns for putting rural schools on a better financial basis appear to have been successful. But all such work has had to be more or less a groping in the dark. Accordingly, light is needed on the possibilities and limitations of methods for securing better rural-school support. The writer, because of his interest in publicity for schools, was asked to start such a study. This analysis is the result.[1]

A good publicity man attacking the problem of how to secure better support for rural schools through publicity would at once set about securing accurate information on the following: (1) Who are the persons that actually decide the amount of support for rural schools? (2) What are their chief characteristics that must be considered if such support is to be secured? (3) What are the direct appeals that will influence them favorably, and how may their objections be handled? (4) How may these appeals be presented successfully? (5) Good examples and descriptions of successful appeals or campaigns for better school support in rural communities. (6) A bibliography or list of sources and suggestions for material and its presentation. Let us then consider these in order.

I. Who Are the People that Actually Decide the Amount of Support for Rural Schools?

One of the axioms of an effective publicity man is: There is no such thing as a general public. To him the public always means certain well-defined groups of varying importance that must be approached by different ways and with different

[1] This article is adapted from a paper prepared for the Committee on Rural Education of the National Council of Education, N. E. A. The paper would have been impossible save for the cooperation of the rural education department at Teachers College. With the aid and approval of Professor Mabel Carney, six students in the department, Misses Frances H. Clark, Osta B. Feurt, Esther E. Jastram, Orpha McPherson, Mrs. Katherine M. Steele, and Mr. H. T. Phillips, met the writer once a week for six weeks in the spring of 1922. This study in the main consists of conclusions reached by the group, all of whom had previously had wide experience in the rural education field. The section on motives became possible only through the cooperation of eighty leaders in rural-school work in various parts of the country, whose names and positions cannot be listed here for lack of space.

appeals. His first work in a given situation is always to discover the important groups and their relative importance.

In listing such groups for the support of rural-school work, one thing must be clearly in mind at the outset. If the move for rural-school support involves legislative action by the state—and most effective moves of this kind do—the rural voters control the situation. Whether the increase is to be determined by constitutional amendment, legislative enactment, a general board, the county superintendent, or local voters, in practically every instance the rural voters decide the issue. City voters seldom object to more state money for rural schools; rural voters often do. For instance, several years ago in Wisconsin the writer found a county without a second supervising teacher because the farmers feared the increased expense. This expense would have been paid by the state and that county's part would have amounted to $33 a year or one-half cent per pupil in the county system. In another county for the same reason a supervising teacher was refused a salary raise of $100 per year when that county's increase for the raise would have been less than one dollar. From a political standpoint, if enough rural voters really desire increased state school revenue, governing officials and voters in other districts will not seriously oppose them.

For all rural situations certain distinctions as to people should be made. The following classifications are suggested, each group of people being named in italics. It will be understood that the classifications are not always mutually exclusive. In most rural areas, *men* have much more to say about increases in school revenues than do *women*. *Children* have much less influence on the decision than in cities because they include few of high-school age and are consequently less considered. Moreover, they are not aroused to the need of more money for schools by their teachers, as are city children under older and more experienced teachers and superintendents. Fortunate is the rural area with a fair number of *moderns*, the younger voters, those who have been away to school and those who have fairly recently moved into the district. Many of the most backward rural

regions are afflicted with many *old fogies,* those who have never been where they could get new ideas and have never secured such ideas through reading. They are often the descendants of families that for several generations have been living in the vicinity. *People with children and those without children* constitute two entirely different groups that must be met with altogether different appeals. *Individuals as distinguished from organizations* are much more important among farmers than in cities. But farmers can still be reached through their organizations if their leaders are convinced. A serious difficulty arises here, as compared with city conditions, because the organizations of farmers are seldom coterminous with the school fiscal areas. However, in many small districts there are factions that will vote as the leaders do, and foreign language groups are easily reached if their leaders are converted.

The following distinctions should be made in certain localities. *Owners* may live on their farms and possibly have children in rural schools, or they may live at a distance. In the latter case their own children or the children in whom they are interested may attend school in districts not including the farm property. Owners are markedly different from *tenants* who may be temporary and who in any event are much less important for deciding increases in rural-school taxes. The inhabitants of certain small rural areas may be largely one *foreign language group.* Small areas in which different languages are strongly represented may easily prevent moves for larger school-support districts if the matter is left to local voters. *Whites and negroes* are distinguished by the laws requiring separate schools for the two races in the southern states. The distinction arises anywhere if negro children become sufficiently numerous; but in the northern states few negroes settle in rural areas. *Private schools* in rural areas usually exist mainly for secondary-school work; *parochial schools* for elementary schooling. The latter sometimes occur in groups using foreign languages.

II. Some Characteristics of Those Who Decide the Amount of Support for Rural Schools

The list here given aims to include those characteristics normally to be expected. In certain places, the opposite or at any rate marked variations, may be found on certain items. It is probably safe to expect any rural population to exhibit nearly all the characteristics here given until evidence to the contrary has been secured. *(1) Conservatism* sometimes appearing as actual hostility to new ideas. Only a few rural populations fail to exhibit this conservatism—for example, those where nearly all the people have moved in recently (as in an irrigated area), or where a land company is trying to attract settlers or outside purchasers, or where there is an economic crisis as there is at present in the west central states. Yet even in such cases the reactions are often towards conservatism. *(2) Backwardness.* In many states the rural population does not contain the most intelligent, the best informed, the most literate, or the most progressive groups of the population. This means that the rural areas in these states are likely to have that portion of the state's population which knows least about schools and is least qualified to pass upon questions connected with them. *(3) Intense individualism.* Land ownership, personal feuds, and the natural isolation of farm life produce less following of leaders than in cities, and a much greater clinging to local "rights." *(4) Difficulty in understanding school financing on a large or modern scale* is experienced by many farmers for very natural reasons. Except in dairy regions, many of them use comparatively little cash. They cannot easily conceal their property from taxation and taxes must be paid in cash. It is a commonplace among politicians that the rural vote can always be secured on an economy platform. Part of the trouble in connection with school improvement may be due to a lack of knowledge of ways of appealing to farmers. Some of them seem to have plenty of cash for the mail-order houses of which they are the largest customers. *(5) Difficulty in seeing the economic value of education.* Farmers naturally see little relation between the schooling given in their

schools by cheap, young, and incompetent teachers and earning capacity on the farm. The value of high-school education as opening a way to the professions appeals little to many farmers because it means taking their children away from home and neighborhood. *(6) Inaccessibility through reading.* The rural population is relatively inaccessible through printed matter. It must be reached by word of mouth and by personal contacts to a greater extent than is the case with city people. But fortunately the rural people probably read more carefully and think more about what they read, than do city people. Furthermore in rural districts, a school circular will encounter far less competition from printed matter than in cities.

III. Successful Appeals to Rural Voters for Better School Support

A good publicity man endeavoring to secure better rural-school support would prepare both for direct appeals and for handling objections. As regards direct appeals, the common sense way is, of course, to appeal to the strongest motives that will induce farmers to vote additional school taxes. But since no usable knowledge of these motives can be found, the writer has endeavored to start work that would supply this lack. The group of students referred to above formulated all the important motives they could think of in this connection, nineteen in number. Great pains were taken to phrase these motives in simple language free from practical ambiguities but no effort was made to split psychological or other theoretical hairs.

The next problem was to decide which motives influence voters. It is comparatively easy to find out how people vote, but it is a wholly different matter to discover just why they vote as they do. Probably in many cases they themselves do not know. The only feasible method seemed to be to take the judgment of experienced leaders in rural-school work. While it was felt that the results of this method could not be absolutely certain, it seemed reasonable to expect that they would be far more

valuable than any mere analysis and theorizing. Accordingly, the list was then submitted to eighty-six men and women who had had successful experience in getting more money for rural schools, on at least a rural high-school or county-wide basis.

On the basis of his own experience and observation, each of these rural-school leaders signified by ranks from 1 to 19, "the relative strengths of the motives that, by and large, influence people to vote more support for rural schools." The rankings were then combined by the simple method of adding the rank numbers for the various motives, the one with the smallest sum of rank numbers being considered the strongest, and so on. Since twenty-six of those ranking the motives had had at least state-wide experience, their rankings were put into a separate group.[2] The sixty with somewhat smaller experience were kept in another group.[3]

The two different combinations are given in Lists I and II. It will be noted that the motives do not have precisely the same ratings in the two lists, although the agreement is noticeably high. Just why these differences obtain, no one at present knows. In all probability, the state workers are accustomed to working on a larger scale with larger projects, where there is more hope of success. They doubtless have more frequent contact with progressive citizens and think more often of securing the support of city people and members of the legislature or of congress. It is possible also that the local list would have given different results with a different proportion of Wisconsin people.

It cannot be too emphatically stated that these two formulations are far from final or authoritative. They do, however, represent the combined judgment of authorities in this field

[2] The twenty-six persons of at least state-wide experience in rural education have one or more representatives with close associations with rural-school work in each of the following states: Alabama, Arkansas, Connecticut, Delaware, Indiana, Iowa, Kentucky, Maine, Maryland, Michigan, Mississippi, Missouri, Montana, Nebraska, New York, South Dakota, Tennessee, Texas, Virginia, and Wisconsin. Several of these people are national authorities in the field, with approximately nation-wide observations.

[3] Twenty-five of these were from Wisconsin and the others represented experience in Alabama, Florida, Iowa, Kansas, Kentucky, Maine, Maryland, Michigan, Missouri, Nebraska, New Jersey, New York, North Carolina, Pennsylvania, Tennessee, Texas, and Virginia.

and, as such, are undoubtedly apt to be nearer the true rankings than the isolated judgments of individuals can be. It is recommended that any person who wishes to undertake publicity work for better school support should select the list that better fits his particular situation. He should then run down the list, checking each motive that applies. Not all motives can apply to every situation. For instance, the motives relating to high schools clearly cannot be utilized in a situation where a high school is not even mentioned. Of the motives that do apply, he can be reasonably certain that any one is stronger than those coming after it and weaker than those coming before, unless he has unimpeachable evidence to the contrary. Just how much stronger or weaker it is, no one can now say. It may be that some of the motives coming close together are virtually ties. The writer is continuing this work on motives and hopes to present at a later date evidence on their approximate relative strengths. After such strengths are roughly determined, all publicity work should stress the motives in the order of their importance.

CHECK LIST I. RELATIVE STRENGTHS OF THE MOTIVES THAT INFLUENCE PEOPLE TO VOTE MORE SUPPORT FOR RURAL SCHOOLS

(As determined by the rankings of 60 successful school workers of at least rural high-school or county-wide experience in 17 states.)

1. To give one's children as good a chance in life as any children get.
2. To give children a better education than their parents had.
3. By making farm life more attractive through good schools children may be induced to stay on the farm.
4. To give children an easier time than their parents had.
5. To give children something of value which cannot be taken from them.
6. To enable the community to have as good a school as some other particular school.
7. Good schools will bring a better class of farmers to the neighborhood and so give one's children better associates.
8. In the case of a rural high school, children can go to it from home, thus escaping possible moral dangers of boarding in town.
9. Farm property will increase in value because of good schools.
10. Increases will enable the district to have the best school in its class or region.
11. By having a good rural high school, children can stay at home longer and be a greater comfort to their parents.

12. The increase is not so heavy as it appears at first thought.
13. The increase will bring much more state school aid of which the cities will pay the bulk.
14. The results claimed to be possible under the increases have actually been secured in real schools by similar expenditures and these schools can be visited and observed by the farmers or their representatives.
15. Educational authorities, such as members of the State Department, the State University, a Normal School, or the County Superintendent, say that these amounts of expenditure are needed.
16. To give the children of other people in the country as good a chance as the city children have.
17. A community center for neighborhood enjoyment can be created by spending more money, especially through a consolidation project.
18. Expenditures for public schools are the one public expenditure that cannot be withheld at one time and later made up; a citizen voting down such expenditures does so at his own and the country's peril.
19. It is good politics always to support increased expenditures for rural schools, when such are shown to be needed and when there is reasonable assurance that the money will be economically used.

CHECK LIST II. RELATIVE STRENGTHS OF THE MOTIVES THAT INFLUENCE PEOPLE TO VOTE MORE SUPPORT FOR RURAL SCHOOLS

(As determined by the rankings of 26 successful rural-school workers of at least state-wide experience in 17 states and familiar with rural conditions in many other states.)

1. To give one's children as good a chance in life as any children get.
2. In the case of a rural high school, children can go to it from home, thus avoiding possible moral dangers of boarding in town.
3. By having a good rural high school, children can stay at home longer and be a greater comfort to their parents.
4. To give children a better education than their parents had.
5. To give children something of value which cannot be taken from them.
6. By making farm life more attractive through good schools, children may be induced to stay on the farm.
7. Good schools will bring a better class of farmers to the neighborhood and so give one's children better associates.
8. To give the children of other people in the country as good a chance as city children have.
9. Farm property will increase in value because of good schools.
10. To give children an easier time than their parents had.
11. A community center for neighborhood enjoyment can be created by spending more money, especially through a consolidation project.

12. To enable the community to have as good a school as some other particular school.
13. The increases will enable the district to have the best school in its class or region.
14. The results claimed to be possible under the increases have actually been secured in real schools by similar expenditures and these schools can be visited and observed by the farmers or their representatives.
15. The increase will bring much more state aid of which the cities often pay the bulk.
16. The increase is not so heavy as it appears at first thought.
17. Educational authorities, such as members of the State Department, the State University, a Normal School, or the County Superintendent, say that these amounts of expenditure are needed.
18. Expenditures for public schools are the one public expenditure that cannot be withheld at one time and later made up; a citizen voting down such expenditures, does so at his own and the country's peril.
19. It is good politics always to support increased expenditures for rural schools, when such are shown to be needed and when there is reasonable assurance that the money will be economically used.

All experience in city campaigns for increased school support and in commercial salesmanship as well, suggests the need of determining for any rural situation, the probable objections to the proposed increase. Then offsets to these objections should be ready. In salesmanship an objection is counted a sign of interest. The salesman's business is to transform the energy signified by the objection so that it flows in the direction of his proposition rather than against it. Lack of space forbids the presentation of the complete list of common objections and their offsets formulated by the group. The following samples will give the idea: (1) *Objection.* The old rural school was good enough for my father and for me, so it is good enough for my children. *Offset.* Have other farmers meet this by calling attention to the changes in farm machinery, roads, autos, etc., and ask if the children are not likewise entitled to better things. (2) *Objection.* Let those wanting high-school education for rural children pay for it themselves. *Offset.* Show that this is un-American and appeal for a square deal and an equal chance for the country child with the city child.

IV. How to Present the Appeals Successfully

Successful publicity work here would involve a check list of things to be done, in chronological order for the best results, and a list of cautions to be observed throughout the work.

CHECK LIST OF THINGS TO BE DONE

1. For a particular situation, determine exactly who will decide the amount of support for the rural schools concerned, using the list of groups given in Section I.
2. List these groups in order of importance for their ability to secure the desired increase in school support.
3. Get clearly in mind the characteristics of each important group that must be taken into account if increased school support is to be secured, using the notes in Section II.
4. Select the better of the lists of Motives in Section III, taking the ones that will apply to the situation in the order of their strengths, and decide just which of these will apply to the various groups or individuals to be influenced favorably. Remember that these motives apply to the rural population in general. They may not apply without change to a particular situation and hardly universally to all the groups in that situation.
5. It is probably well to count on using most of the motives sometime so as to be sure of reaching all possible voters.
6. List the principal objections likely to be raised by the various groups, with the probably successful offsets.
7. Decide whether it is better to make an open campaign or to have things settled by a quiet conference with a few leaders. The latter procedure may work easily in small matters or with something that the people cannot take back (as for instance a building that once erected is there to stay); but in the long run frank and open publicity is preferable for most things.
8. Expect that in general it will take a long time to secure favorable action on any marked changes in organization or increases in rural-school support. It is best to plan for five- and ten-year periods at least.

CAUTIONS TO BE OBSERVED THROUGHOUT THE PUBLICITY WORK

1. Much personal work is advisable, especially for securing the early support of leaders and for avoiding unnecessary antagonisms.
2. Do not expect too much from reading matter.
3. Create the desire for better schools and discuss increases in costs later. Farmers in private life spend their money, and in public affairs vote taxes, for anything they desire strongly enough.
4. The material intended for any group must of course be adapted to that particular group. Employ illustrations and language that will appeal to its members, and let these be presented by those personally acceptable to the group.

5. Farmers care little or nothing for per capita costs. They wish to know what the improvement will cost the district as a whole or them individually.
6. In many places it will be necessary to show clearly that there is no hope of materially bettering the rural schools until a larger administrative unit, such as the consolidated district or the county unit, is secured.

V AND VI. EXAMPLES OF SUCCESSFUL APPEALS FOR BETTER RURAL-SCHOOL SUPPORT; BIBLIOGRAPHY

Scattered examples of material and some references have been found. But in the time available we have not been able to analyze or test them so as to permit a worth-while estimate of their value. Until this is done, they cannot be thrown into usable form.

Meanwhile, those who must issue publicity material for better rural-school support, will find the quickest sources of supply to be: *The Journal of the National Education Association*, Washington, D. C.; the Educational Division of the International Harvester Company, Chicago; and *School Life*, which can be secured through the United States Bureau of Education. What appears to be valuable material in this field has been sent out in connection with campaigns from the state departments of Alabama, Indiana, Maryland, Missouri, South Dakota, Texas, and Utah. The publicity Committee of the Wisconsin State Teachers' Association has issued many appeals directed especially toward the improvement of rural schools. Very promising material in printed form has come from Montgomery County, Alabama, from Monterey County, California, and from Cuyahoga County, Ohio. Similar material in mimeographed form has been issued by Prince Georges County, Maryland.

Many valuable suggestions and references on general school publicity in fiscal matters will be found in Alexander's "School Statistics and Publicity"[4] and in "Publicity Campaigns for Better School Support" by Alexander and Theisen.[5]

[4] Alexander, Carter. *School statistics and publicity.* New York: Silver Burdette and Company, 1919. 332 pp.

[5] Alexander, Carter, and Theisen, W. W. *Publicity campaigns for better school support*, (School Efficiency Monographs). Yonkers: The World Book Company, 1921. 164 pp.

In the foregoing the writer has merely tried to start work which might ultimately enable rural-school leaders to use publicity methods for better school support as effectively as do leaders in other fields. No one realizes better than he how unsatisfactory is our present knowledge of such matters. He has found it extremely difficult to discover sound methods of procedure in publicity work; but that is no reason for refusing to make the attempt. He will, of course, be gratified if the analyses and suggestions here given are found helpful by rural education leaders, or if they stimulate such leaders or educational research men to produce things that will be helpful. But he is much more concerned with securing knowledge that is accurate and reliable, with finding out just where these suggestions are sound and where they are weak, and just how sound or weak they are. We need something better and more dependable as soon as possible.

If the rural-school leaders can find out how to use publicity methods effectively, they will be greatly relieved from their greatest source of worry at present—inadequate financial support. But guesses, half-baked ideas, or biases about publicity methods cannot set them free. Only the truth—accurate knowledge of such methods—can do that. Hard work by many capable persons will be necessary before this knowledge can be secured. But the work will offer a highly stimulating challenge to competent educational research workers, particularly those well grounded in social psychology and in accurate observation of practical politics. Any such research worker who wishes to make his activities count for the improvement of rural education can be sure of one thing. Research by him on rural-school work along the lines pursued so far will in most instances produce only trivial results compared to what the same energy could accomplish if it were efficiently applied to securing accurate knowledge of how to get better rural-school support through publicity work.

THE STANDARDIZATION OF TESTS AND SCALES.[1]

PETER SANDIFORD

Ontario College of Education, University of Toronto

The problems of the standardization of tests and scales may be considered from many points of view, chief of which are the following: (1) the units and the zero points of tests; (2) the criteria of standardization of tests; (3) the validity of tests. Although these three aspects of the problem are not mutually exclusive, the amount of overlapping is not sufficiently serious to embarrass a writer who attempts to discuss them. As the discussion of (1) and (2) demands a knowledge of the mathematical theory of statistics to which I cannot lay claim, the major part of this paper will centre around (3).

UNITS AND ZERO POINTS

In all physical standards there are zero or reference points and the units are arranged so that all parts of the scale are strictly comparable. Most of the traits, however, that psychologists and teachers wish to measure are variable ones, and the task of discovering suitable zeros and units for them is far from easy. At present the utmost confusion prevails owing to the variety of units and zeros in use. There is a different zero point for practically every scale and test, and the units used are as numerous as the zeros. In the Binet-Simon scale the zero is the time of birth, and the unit one year of mental growth. The zero is one that everybody understands, but the unit probably varies from year to year. It is extremely unlikely, for example, that the amount of mental growth between 5 and 6 is the same as that between 15 and 16. The unit, therefore, is an unsatisfactory one.

I am in no position to dogmatize on the subject, but it seems to me that the science of mental measurements will be furthered

[1] A paper presented at the joint meeting of Sections I and Q of the American Association for the Advancement of Science, Toronto, 1921.

most assuredly by the adoption of a reference point in the center of the scale with a zero having some definite relationship to it, and of units which take into account the variability of the measurements from the central tendency. The index method of Weiss and the "standard measures"—"s", "m" and "v"—of Kelley were early attempts at the solution of the zero and unit problem. McCall[2] has more recently recommended the mean performance of children between the ages of 12 and 13 as a reference point (50) with a zero point situated 5 S. D. below it, and a "T" as a unit of measurement. The unit "T", named in honour of Thorndike and Terman, is the standard deviation of the variability of twelve-year-old pupils multiplied by ten. For this unit McCall claims the following advantages. It is "equal at all points on the difficulty scale, which is the chief characteristic of the unit employed by Thorndike and his students (i. e., some function of variability as a unit). It is based upon chronological age, which is the chief characteristic of the work of Terman and his predecessors. It is a function of the variability of a defined group and a group which is easily located. A scale which uses this unit reaches as low and as high as the ordinary requirements of practical measurement. Special extension at the top or bottom is a simple process. The subsequent improvement of scale values is simpler still. And not of least importance is the fact that the construction of the scale which employs this unit is not particularly laborious. In sum, the proposed unit combines most of the virtues and eliminates most of the defects of the two chief contemporary methods of constructing mental scales. In a certain sense it unites the two great lines of scale development." Happy indeed will mental testers be if McCall's claims are substantiated. But, seriously, there is so much confusion reigning in regard to zeros and units that it behooves all who have the necessary knowledge and skill to give themselves to the task of helping to create order out of the present chaos.

[2] McCall, William C. "Proposed uniform method of scale construction," *Teachers College Record*, 22:31-51, January, 1921.

The Criteria of Standardization

It is of the utmost importance for the user of scales and tests to know when they have, or have not, been properly standardized. Some of them have been finally standardized before they reach the user; others are sent out tentatively standardized and the records made from them are used in making the norms more perfect. Thorndike's Handwriting Scale, Ayres' Spelling Scale, Pintner-Paterson's Performance Tests, and Terman's Revision of the Binet-Simon Scale are examples of the first type. Thorndike's Supplementary Matriculation Test and the Otis Group Intelligence Test are examples of the second. In the last analysis a test is standardized when additional results or records obtained from its use do not alter the general form of the distributions (measured by variability) or disturb their central tendencies. Theoretically, a test should only be regarded as standardized when an infinite number of unselected subjects have been tested with it and the results statistically analyzed. Practically, however, constancy of norms is reached fairly soon. As we shall see later the chief difficulty in standardization is not in the number of cases, but in the inadvertent use of selected subjects.

There are many minor criteria of standardization which psychologists and educators habitually use, and common sense as well as experience justifies them. The most important one is the assumption that the records of measurement of any trait, providing a sufficiently large number of unselected but homogeneous individuals is taken, will be distributed according to the normal surface of frequency. If the records from a test when graphed show marked deviations from the normal surface of frequency they should be regarded with suspicion. Skewness, however, is sometimes the normal condition of affairs. The weights of adult Anglo-Saxons are skewed to the right; there are more of them over-weight than under-weight. The heights and intelligences of these same Anglo-Saxons are, however, normally distributed. Terman was perfectly correct in expressing satisfaction when the I.Q.'s of his subjects were found to be distributed according to the normal surface of frequency.

A second important criterion, which is used where the records show a steady increase or decrease, is the regularity of the graphed result. If there are "jogs" in the curve, sudden changes in direction, they are justly regarded with suspicion. The chances are overwhelming against an irregular graph representing a true condition. It is reported of Quetelet that when he first graphed his results for the heights and weights of boys and girls at successive ages, he was greatly perturbed at the irregularity of his curves around the age of puberty, particularly when he discovered that the girls for about three years exceeded the boys both in height and weight. According to the story, he then did a very unscientific thing. He cooked his results by the addition of the records of as many tall and heavy boys, and as many short and light girls as he could find. The tale is probably apocryphal but it illustrates my point. Quetelet's feeling that the curves should be regular was a sound one. Where he went astray was in faking a result; he should have remained content when additional results of unselected cases did not modify the norms of his earlier records.

A glance at the graphical representation of norms in almost any magazine is sufficient to cast doubts on the accuracy of the standardizations. For example, J. E. Wallace Wallin[3] in an article entiled "Age Norms of Psycho-Motor Capacity" gives some graphs, the irregularities of which are almost certainly due to imperfect standardization. The addition of further results would undoubtedly lead to a smoothing of the curves.

Some years ago I gave T. A. McWhorter, one of our graduate students in education, the task of further standardizing some of the performance tests described in Pintner and Paterson's "A Scale of Performance Tests." Space forbids me giving all the new standards. Table I and Figure 1, drawn from it, will suffice for illustration. Figure 1 illustrates the changes in the median curve caused by the additional cases. There is no doubt at all which is the better.

[3] *Journal of Educational Psychology*, 7:17-24, January, 1916.

TABLE I. FURTHER STANDARDIZATION OF THE HEALY PUZZLE "A." (TIME UNITS)

Age	Pintner		Pintner and McWhorter	
	Totals	Medians	Totals	Medians
4	5	D. N. C.		
5	43	D. N. C.	67	D. N. C.
6	108	D. N. C.	183	D. N. C.
7	138	131	250	185
8	147	117	275	118
9	122	86	209	83
10	117	70	208	60
11	105	54	193	55
12	88	46	179	52
13	56	38	143	43
14	44	55	108	46
15	25	30	46	30
16	7	35		
Total	1005		1861	

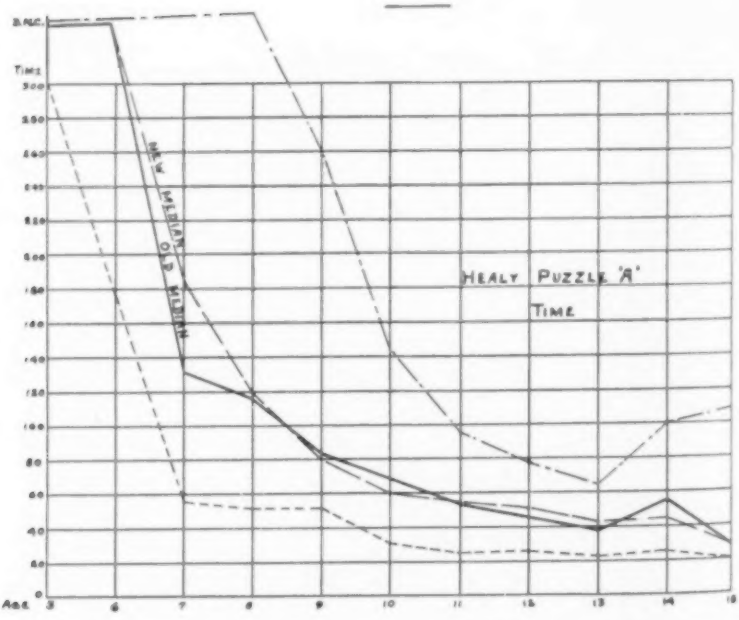


FIGURE 1. MCWHORTER'S FURTHER STANDARDIZATION OF PINTNER AND PATERSON'S RESULTS

The Validity of Tests

Any one who has carefully studied the results of psychological examining in the United States Army[4] will have been struck by the great diversity of the intelligence ratings there shown. Not only do the intelligence ratings differ greatly according to the camps in which they were made (Part III, chapter 10), but they also differ according to occupational groups (chap. 15), according to the arm of service (chap. 16), according to nativity (chap. 6), according to race (chap. 8), by states (chap. 5), and in every conceivable way. Psychologists have known these facts in a general kind of way for many years, but the findings of such wholesale testings have driven the facts home as never before. Even the man in the street has had his interest aroused. It is no part of my task to present these materials again. Anybody can study them for himself, but I would like to present two tables and a diagram in order that my subsequent argument may be better followed.

The editor of the Memoir gives tables of distribution by states but does not analyze them. I have taken the trouble to summarize Table 200 which gives the alpha scores by states of Groups I and II, white draft. These men took only the alpha test. They represent, therefore, the literate and more intelligent men found in the American army. Table II gives a summary of the findings. The table is arranged in the order of least frequency of grade D.

Utah is seen to head the list, but if all grades are taken into consideration, Washington, with the exception of the special case of the District of Columbia, can claim to be the most intelligent state in the Union, closely followed by Oregon, California, Idaho, Wyoming, Connecticut, Massachusetts, and Vermont. Practically all the Northern States are above the grade for "Total White Draft" and practically all the Southern States are below. New Jersey, at the bottom of the list for Grade D, is

[4] Yerkes, Robert M. Ed. *Psychological Examining in the United States Army.* (Memoirs of the National Academy of Sciences, Vol. 15.) Washington, D. C.: Government Printing Office, 1921. 890 pp.

TABLE II. PERCENTAGE DISTRIBUTION OF LETTER GRADES IN INTELLIGENCE BY STATES

States	D	C	B	A
Utah	.7	79.4	13.8	6.1
California	1.1	74.8	15.7	8.2
Washington	1.6	71.7	15.6	11.1
Oregon	2.0	71.6	16.7	9.7
Nevada	2.4	79.0	14.9	3.7
Idaho	2.8	74.0	16.3	6.9
Wyoming	3.0	74.1	16.3	6.6
Connecticut	3.0	74.5	14.3	8.1
New York	3.2	79.5	11.3	6.1
Vermont	3.4	74.6	12.8	9.2
Massachusetts	3.9	73.0	13.0	10.2
Maine	4.37	76.5	11.0	8.3
Montana	4.4	78.7	12.2	4.7
Minnesota	5.0	78.3	10.9	5.8
Nebraska	6.0	78.1	11.9	4.0
District of Columbia	6.5	62.4	15.6	15.6
Iowa	7.0	75.0	11.3	6.7
Colorado	7.1	68.4	14.2	10.1
Kansas	7.9	76.9	10.2	4.8
Pennsylvania	8.3	74.0	11.4	6.3
New Hampshire	8.7	71.6	14.8	4.9
Rhode Island	8.7	78.0	9.3	4.1
South Dakota	9.0	75.6	9.2	6.2
New Mexico	9.2	79.4	7.6	3.8
Michigan	9.3	74.2	11.0	5.7
Ohio	9.3	68.5	13.7	8.5
Illinois	9.7	73.0	11.4	5.7
Missouri	9.8	75.7	9.9	4.6
Total White	10.0	73.7	10.7	5.6
Indiana	10.3	77.6	9.5	2.6
Florida	10.7	70.4	6.4	12.8
Oklahoma	11.3	77.8	7.7	3.2
Texas	12.2	78.2	6.9	2.6
North Dakota	12.6	73.3	9.6	4.6
West Virginia	15.1	71.2	9.4	4.3
Wisconsin	15.4	68.8	10.1	5.7
Maryland	15.6	73.1	7.3	4.1
South Carolina	16.7	76.1	5.0	2.2
Louisiana	19.1	73.3	5.4	2.2
Alabama	19.2	71.6	6.0	3.2
Mississippi	19.3	75.8	3.9	1.0
Virginia	20.0	66.2	10.3	3.6
Tennessee	20.2	69.3	7.9	2.6
Georgia	20.8	70.4	6.7	2.1
North Carolina	20.9	72.5	4.4	2.1
Arkansas	23.8	72.3	3.2	.6
Delaware	26.8	61.2	5.6	6.4
Kentucky	28.5	64.7	4.6	2.3
New Jersey	28.7	56.3	9.6	5.4

a notable exception. Kentucky and Arkansas run a close race for bottom place. Too much stress, however, must not be placed on the table as a reliable guide to the relative intelligences of the various states.

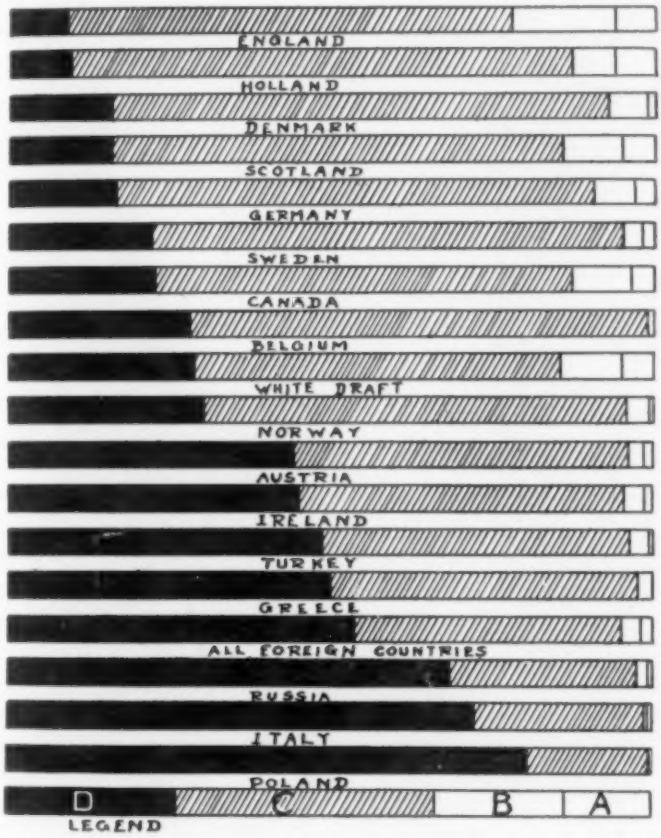


FIGURE 2. INTELLIGENCE BY NATIVITY OF FOREIGN-BORN MEN. GROUPS I, II, III, WHITE DRAFT

Now consider Figure 2, taken from the Memoirs.[5] The intelligence of the recruits born in England is seen to be considerably greater than that of the White Draft, while that of Scot-

[5] *Ibid.* p. 295.

ish and Canadian recruits is also very high. Poles, Italians, and Russians make a very bad showing.

In the combined scale, as shown by Table III, the order is somewhat changed. Canada now heads the list and Italy foots

TABLE III.

Country	Mean	Standard Deviation	Number of Cases	Mean Mental Age
Canada	13.74	2.71	948	13.29
England, Ireland, and Scotland	13.37	2.83	1214	13.00
Denmark, Norway, and Sweden	13.30	2.41	1610	12.95
Germany and Austria	13.17	2.69	597	12.85
Greece	11.90	2.57	573	11.86
Russia	11.16	2.96	2701	11.28
Italy	11.04	2.598	4002	11.19

it. The southern Europeans compare badly with northern Europeans and Canadians.

Now what is the moral of all this? If a test is standardized on the Pacific Slope, as Terman's Revision was, it will probably include more than a just proportion of the more intelligent children. It will, therefore, set too high a standard for the United States as a whole and to that extent will be imperfectly standardized. Similarly if the standardization is made in a community containing a large proportion of English and Canadian immigrants the same result will follow. Conversely, if it is made in the southern States or in a community with a large proportion of southern Europeans, the standardization will rate too low. Any test, therefore, reputedly standardized for the United States will in future have to contain results from north and south, east and west.

If negroes are to be considered then the problem becomes still more complicated. Table IV shows the findings of the army tests on racial intelligence.

The negroes show a striking inferiority in intelligence when compared with whites. Further, the negroes of the Northern States are markedly superior to those of the Southern States.

Turning now to Canada, I have to confess at the beginning that we are practically dependent upon the United States for our

TABLE IV. INTELLIGENCE OF WHITES AND NEGROES

Race	Number of Cases	Percent Making Grade						
		D−	D	C−	C	C+	B	A
Whites, Groups I, II, IV.....	93,973	7.0	17.1	23.8	25.0	15.0	8.0	4.1
Negroes, Group IV	18,819	49.0	29.7	12.9	5.7	2.0	0.6	0.1

various tests and scales. While we are grateful for the hearty way in which our fellow-educators cooperate with us and render every assistance, some of us wonder if tests standardized in the United States are valid for Canada. I have not many results to place before you, but such as I have, lead to the conclusion that either we must make tests and scales of our own or re-standardize yours.

Dr. Hamilton, a graduate student of the University of Toronto, constructed three handwriting scales to measure the handwriting of Ontario school children. This piece of work was peculiarly appropriate since all Ontario children are taught to write in the same way from the same copy books. It is obvious, however, that there cannot be a single standardization of handwriting that will fit all cases.

A. W. R. Doan, another of our graduate students, in measuring the school buildings of Toronto, was forced to modify somewhat the Strayer scale which is so largely used in the States. I cannot conceive that any one scale for school buildings can be used, say, to measure the school buildings of Northern New York with its severe winter and the school buildings of Florida where winter is unknown.

C. Sansom[6] used the Otis tests in the Highland Practice School, Edmonton. Since his results have been reported it will be sufficient for me to allude to the fact that according to his results these Canadian pupils were somewhat below those of the United States. The cases, however, were too few to justify any generalization from them.

[6] Sansom, C. "An analysis of scores, Otis Scale," *The School*, 9:635-45, May, 1921.

During the past two years Professor Bott and I have given the Thorndike Columbia test to various groups of students in the University of Toronto, and to high-school pupils of Toronto about to try the matriculation examination. The results for the university students are given in Table V.

TABLE V. UNIVERSITY OF TORONTO SUMMARY OF THORNDIKE TESTS, 1920

Score	Second Year Medical Students (Men)	New Graduates Ontario College of Education			First Year St. Michael's			Total
		Men	Women	Men & Women	Men	Women	Men & Women	
110–up	1	1		1				2
105–109.9	1	0		0				1
100–104.9	2	2	1	3				5
95–99.9	3	3	3	6				9
90–94.9	4	3	3	6				10
85–89.9	8	6	5	11	4	2	6	25
80–84.9	2	4	8	12	3	1	4	18
75–79.9	5	1	7	8	9	3	12	25
70–74.9	0	2	7	9	9	7	16	25
65–69.9	5	1	4	5	10	7	17	27
60–64.9	2	2	2	4	6	4	10	16
55–59.9	1				4	4	8	9
50–54.9	1				6	1	7	8
45–49.9	1				1		1	2
40–44.9	1				3		3	4
35–39.9					1		1	1
Total	37	25	40	65	56	29	85	187
Median	85.2	88.0	79.8	83.4	67.6	67.2	67.6	

A comparison of our results with those of Thorndike is presented in Table VI. Some results for the Toronto high-school pupils are also included.

It is difficult to draw conclusions from these results, since the Toronto students include both graduates and women. Thorndike thinks the women are penalized to the extent of five points by the form of the examination. If we set the low scores of the women against the high scores of the graduates we find that the Toronto students rate about equal to those of Columbia, Brown, and Rutgers. The high-school pupils were taken without selection from those proposing to try matriculation the following month. The presence of girls and of those who subsequently failed in matriculation would naturally lower the scores.

TABLE VI. PERCENTAGE DISTRIBUTION OF STUDENTS ACCORDING TO THEIR SCORES ON THE THORNDIKE TEST

Score	Students at Brown, Columbia, and Rutgers (Thorndike, 500 cases)	Toronto Students (187 cases)	Toronto High School Pupils (175 cases)	Second-rate Colleges and Agricultural Schools (Thorndike, 500 cases)
10–19				1.5
20–29				6.0
30–39	1	0.5	1.1	14.5
40–49	3	3.2	6.9	24.5
50–59	9	9.1	14.3	24.0
60–69	22	23.0	25.7	18.5
70–79	26	26.8	28.0	8.0
80–89	21	23 0	16.6	2.0
90–99	13	10.2	5.7	0.5
100-up	5	4.3	1.7	
	100	100.1	100.0	99.5

This year Professor Bridges, Professor Bott, and I have given a modified alpha examination to several hundred students in the University of Toronto. The results in comparison with those obtained at Ohio State University are summarized in Table VII.

TABLE VII. MODIFIED ALPHA EXAMINATION RESULTS AT OHIO STATE UNIVERSITY AND UNIVERSITY OF TORONTO

	Medians	A. D.'s	Number of cases
OHIO STATE UNIVERSITY			
1st year students	102	24.3	2021
2nd year students	122	24.6	161
3rd year students	124	24.1	115
4th year students	135	23.2	723
Graduates	133	21.8	51
UNIVERSITY OF TORONTO, MEDICAL			
1st year students	112	21.8	400
2nd year students	117	16.9	52
3rd year students	118	22.6	46
Graduates	130	21.0	118

In comparison with Bridges' results for Ohio State University, Toronto students on the whole are shown to be slightly inferior. The results obtained by the fourth year students of Ohio State University should be discounted a little since they had worked through the same examination a year previously.

When I offered a paper last May on the standardization of tests, I intended to present a Canadian standardization of the Ayres' Spelling Scale. Delays in printing have prevented me from carrying the work to a conclusion. Only the results of a preliminary survey can be given to-day. They are shown in Table VIII. Canadian pupils apparently spell better than those

TABLE VIII. PRELIMINARY RESULTS OF AYRES' SPELLING TEST IN CANADIAN SCHOOLS

School	Grade	Number of Pupils	Column Used	Percent Correct	Ayres' Standard
I	5	34	P	89.3	79
II	6	30	R	88.	79
	7	35	T	79.1	79
	8	35	U	83.3	84
III	8A	41	U&T	89.8	86
	8B	41	U&T	88.6	86
	7A	41	U&T	80.	76
	7B	38	U&T	82.	76
	7C	36	U&T	87.4	76
IV	8	47	P	96.5	98
	7	113	P	93.6	94
	6	128	P	92.1	88
	5	185	P	77.2	79
	4	138	P	64.1	66
V	8	106	U	86.8	84
	7	104	T	83.2	79
	6	100	S	79.6	73
VI	8	100	U	87.8	84
	6	100	T	78.	66
VII	6	64	R	81.2	79
	8	24	U	79.2	84
VIII	6	123	R	71.2	79
	8	88	U	73.6	84
IX	8	83	U	87.	84
X	8	100	U	86.8	84

of the United States. They ought to, for Canadian schools devote about 50 percent of their time to "tool subjects", chiefly the three R's. As far as one can judge from study courses and time-tables, American schools spend only about 20 percent of their time on tool subjects, the remainder on content subjects. The transition from the older to the newer forms of spellings probably affects adversely the pupils of the United States more than those of Canada.

I am afraid that some of the specific evidence I have presented may not sound very convincing, but it is sufficient to indi-

cate that general standardizations, except for special purposes, are practically useless. What we need are finer and special standardizations. We need them not by countries, states or provinces, but by cities, wards of a city, and for each racial or social group within the wards of a city. Standardizations by grades and by ages are also too coarse. We need them by ages within the grades. And even here different races and different social and occupational groups should be taken into account. I feel convinced that the orgy of indiscriminate testing that we have had recently on this continent has not served our best ends. Especially have we been at fault in regard to interpretation. Tests standardized for one specific purpose in one area have been used uncritically for other purposes in other areas. The sons of college professors have been triumphantly compared with the sons of labourers and negroes. College professors' sons should be compared with college professors' sons, sons of labourers with sons of labourers, and negroes with negroes. The geographical and social environments should always be taken into consideration. A negro of the south may not make as good a score in an intelligence test as a Harvard professor, but as a picker of cotton he will be infinitely superior.

In the same way an unthinking use of American tests in Canadian schools will lead to erroneous conclusions. Some of the tests, such as those in history and geography, we cannot use at all. Others, we are convinced, need re-standardization. Canadian spelling should be contrasted with Canadian spelling, Canadian intelligence with Canadian intelligence, for in spite of the mobility of population on this continent, most Canadians will live and die in Canada. It is on these grounds that in spite of my great affection for the United States and its people, I advocate, in season and out of season, the desirability of standardizing all the tests and scales we use in Canada according to local Canadian standards and not according to any other.

THE INTELLIGENCE OF TEACHERS IN TRAINING, WITH A NOTE ON PRACTICE-EFFECTS WITH INTELLIGENCE TESTS

SAMUEL RENSHAW
Department of Psychology, Western State Normal School, Kalamazoo, Mich.

I

The belief is more or less current that schools and colleges devoted to the training of teachers are recruiting their student bodies from inferior high-school graduates. A few references are cited which are being quoted to sustain this view:

Yoakum and Yerkes (1) report scores made on the Army Alpha Test by normal-school men and women and compare them with scores made by students from colleges, universities, and technical schools. The scores for the teachers in training are invariably reported lower than the others. To quote from their tables (pp. 164-171), the median score for men students from five normal schools is 115 Alpha points. The median for normal-school women is 111. For college men the median is listed at 130, and for college women it is 127.

Thurstone (2) elsewhere prints figures and diagrams showing the inferiority of the median and other percentile scores for normal-school freshmen compared to the scores of freshmen in engineering and liberal arts colleges. These comparisons are based upon scores made on the Thurstone Test IV, Intelligence Examination for High-School Seniors and College Freshmen.

Woodworth (3) makes a comparison of the frequency distributions of an army draftee sampling, which, at least by implication, he regards as typical, with a college freshman sampling. Alpha test scores form the basis of comparison. He places the median of the drafted men at 65, and that of the freshmen at 150.

Repeated references to such statements as the above have caused many persons to be alarmed at the growing "attenuation" of the stock from which public school teachers are being made. It is the purpose of this paper to examine critically this view and

to offer some evidence in contrast to its position by presenting the distribution of the Alpha scores of a principal sample (1199 of a total registration of 1294) from the student body in the second largest of Michigan's normal schools.

Table I shows the distribution of the scores. The test was Form 6 of the Alpha Army Test, administered by the writer to the students in two groups of about 600 each. Women students outnumbered the men by about two to one—a fact worthy of consideration in interpreting the results, since it is generally understood that women tend to make slightly lower scores than men on Alpha.

TABLE I. DISTRIBUTION OF THE ALPHA SCORES. NORMAL-SCHOOL GROUP

Scores	Frequency
190–199	4
180–189	28
170–179	41
160–169	100
150–159	145
140–149	194
130–139	200
120–129	185
110–119	122
100–109	86
90–99	61
80–89	18
70–79	10
60–69	4
50–59	1
Total.....	1199

The true median for this distribution was computed at 135.65; the mean was found to be 134.95; standard deviation, 23.58; mean deviation, 18.79; probable error, 15.90. The probable errors of the standard deviation and of the mean were found to be 0.32 and 0.46 respectively. The customary interpretation of this sort of fact (for example in the case of the P.E. of the mean) would be that the means of similar groups upon repeated remeasuring with the same or an equivalent scale would in half the findings vary not more than 0.46 above or below the obtained

mean. Further reference will be made to this statement in Part II of this paper.

Arps and Noble (4) published the distribution of Alpha scores for 5950 Ohio State University students. Of this number 2545 were freshmen, 1346 sophomores, 1097 juniors, 760 seniors, 139 graduate students and 63 were special students. In our group of 1199 there were 1116 freshmen and sophomores, 36 juniors, 19 seniors, 3 graduate students, and 25 "specials." The small numbers in the upper classes is due to the fact that the institution has only been granting the baccalaureate degree for the past two years. Formerly two-year courses only were offered. The university population thus has a far greater proportion of junior and senior students. This is a factor selective of the better grades of intelligence, and would tend to raise the central tendency of the university percentiles.

Table II shows a comparison of the deciles of the two groups of students.

TABLE II. COMPARISON OF THE DECILES OF THE ALPHA SCORES, OHIO STATE UNIVERSITY AND WESTERN STATE NORMAL

Percentile	O. S. U.	W. S. N.
100	211	198
90	170	165
80	159	154
70	150	147
60	143	141
50	136	135
40	128	128
30	120	122
20	110	115
10	96	103
1	61	76

The difference of the percentile scores in the top and bottom fifths for these two institutions is doubtless due to the size of the samples, there being nearly five times as many university students as normal-school students. Again, the slight differences may be due to the larger proportional number of upperclassmen in the university group. So far as this evidence is concerned, it seems that there would be no valid difference in the relative in-

telligence of the two groups of students if the samples are reduced to the same basis of comparison.

The figures of Yoakum and Yerkes (1) are open to objection on the ground that they were obtained from five of the smaller and younger normal schools. Moreover, except for one Pennsylvania school these institutions were located in the southern and southwestern sections of our country. It is a well-known fact that among all classes of schools and colleges the student selection and standards will vary with the different localities. There may be wide differences even though the space that separates two institutions be no greater than the width of a city street. Again, these samples were taken in *war-time*. There was assuredly a tremendous attenuation during those years felt in every college and normal school in the country. Also, the number of subjects from these schools is extremely small—in some cases less than 100. We are not told what percents of the student bodies was represented, or the proportions of the sexes. Further, normal schools differ in entrance requirements, teaching staff, and physical equipment. In some of them matriculation means meeting the standard college entrance requirements; in others, it has been said, the only stipulation is the physical ability to "climb the hill, sign your name, and pay the fee."

Thurstone's (2) figures were obtained from Pennsylvania and Virginia normal schools. The figures are no doubt valid for those states. But one begins to suspect that there are special selective factors at work. The differences which he reports between normal-school freshmen and arts and engineering freshmen are significant. Would the same conditions obtain for states like Indiana, Wisconsin, Minnesota, Illinois, or Ohio? It is doubtful.

Woodworth (3) places the college freshman median at 150. There are good reasons for believing that this is too high. Freshmen in state universities do not run that high. In 1919 the median score for Ohio State freshmen was 129, the 75-percentile 147, and the 25-percentile 108. For twenty colleges combined comprising 3175 freshmen it was 130.

Is it true that the majority of teachers in training are drawn from the middle and lower thirds of high-school graduates? In reply we quote Table III, from a report by Whipple and Cobb (6) on Alpha scores made by high-school seniors.

TABLE III. MEDIAN ALPHA SCORES OF HIGH-SCHOOL SENIORS, EIGHT MICHIGAN CITIES AND TOWNS

City	Number	Median score
Alma	53	121
Milan	30	121
Mt. Clemens	29	124
Mt. Pleasant	48	127
Wayne	38	110
Ypsilanti	25	127
Battle Creek	155	128
Highland Park	160	122
Total	980	122.5 (Mean)

The average of the medians of the 980 Michigan high-school graduates is 122.5, which corresponds approximately to the 30th-percentile of our normal-school group and to the 34th-percentile of the Ohio State University group.

Madsen reported Iowa high-school seniors to average about 128 points on Alpha, which is the 40th-percentile for our group, or one-tenth below the normal-school average. The distributions of the high-school students' scores are unfortunately not available. From the freshman figures, however, it is evident that the better types of normal schools and colleges are not attracting the weaker high-school students.

II

To what extent is an Alpha score a valid measure of "intelligence"? What is the probable error of a single Alpha score? How much "error" will be introduced by the effects of practice in taking tests of this sort?

The Alpha with its several forms permits study of the effects of practice. It is reasonable to assume that a person who has repeatedly taken tests of this sort will make a higher score than the novice who is being tested the first time. If this difference

is found with individuals it will also be found in the case of groups. It would be manifestly unfair to compare two groups of pupils of like age and "mentality," the first of which, say, has had two or three times as much experience in taking tests as the other. Under such conditions error will be introduced if individuals are to be selected for classification or comparative ranking, or if some method such as "critical scores" is used for a special purpose.

So far as consistency or reliability of test scores is concerned there are several methods in common usage which can be employed for this determination. The Spearman prophecy formula, often referred to as "Brown's formula" (8, 101-102) may be used to compute the coefficient of reliability. This is supposed to tell *how well* the test measures whatever it does measure. It is expressed as

$$\text{reliability} = \frac{nr_1}{n + (1 - n)\, r_1}$$

where n is the number of repetitions of the test and r_1 is its self-correlation. Pressey (9) and Mrs. Burgess (10) have pointed out that in reality it tells how consistently the subjects tend to make the same scores in the succeeding tests of the same sort. It is interesting to see how our data concerning this part of the investigation behave in accord with the Spearman expectation.

Kelley (11) recommends the use of the P.E., and Ruml (12) has presented a rank-tangential coefficient as further methods of expressing the "efficiency" of a test. Validity usually means the degree to which the test correlates with some independent criterion. We are not particularly concerned with that here. We are interested in seeing what happens when the same group of subjects are given several alternative forms of the same test.

If we assume that 600 of the 1199 subjects referred to in the first part of this paper had previously taken the Alpha test two or more times, how far would this amount of practice invalidate the comparison of their scores with those of a similar group serving for the first time as subjects in a mental survey?

Dunlap and Snyder (7) repeated the Alpha four times, using different forms, with 44 Johns Hopkins junior and senior men. We quote from their report:

> The relative rating on the third test is somewhat different from . . . the first. If the test had been used to select the best fourth of the group there would have been little difference in the individuals selected, since 9 of the 11 who were in the first (high) quartile in the first test remained in the first quartile in the third (test) while the two who go back move but to the next quartile . . .
>
> . . . It is a matter of passing interest to note the comparative uniformity in the results of the first and third tests in the first (high) quartile and the contrasting fluctuations of results in the fourth (low) quartile. Were the tests given with a view to eliminating the lowest quartile the first and third tests would thus yield quite different results.
>
> If the men in the lower half of the class had had practice in the test equivalent to that actually obtained by the taking of the test twice, and the men who are in the upper half had not had practice, a single test would have given for the class a relative grading which would have differed markedly from either the first or third of the actual ratings, and which would have been maximally unfair.

The writer repeated this experiment under slightly different conditions. The subjects were 24 sophomore and junior students in a laboratory section in a course in Educational Psychology. Forms 9, 6, and 8 of Alpha were given at intervals of three weeks, with especial care to keep the incentives of the students uniformly high. The scholarship of these students would average around C. There were 5 men and 19 women. One man had taken Alpha previously in the "scrambled" form. The rest were novices. The coefficients of correlation between the three tests were as follows:

Tests	r
1 and 2	0.76
1 and 3	0.78
2 and 3	0.86

The mean scores of the first, second, and third tests were respectively 126.2, 153.2 and 161.7. The gain of the average of test two over the average of test one was 21.4 percent; but test three showed only 5.5 percent gain over test two. The mean of

the third test was 28.1 percent greater than the mean of the first test. Evidently there is a notable practice effect and it seems rapidly to become asymptotic, although we are not just sure as to the form of the practice curve. A small amount of practice with Alpha seems to give a student, therefore, quite a considerable advantage over the novice. Suppose, for example, a student makes a score of 120 on his first test. If this also gives him a 25 percent practice-effect advantage, the second time he is tested his score should be 150. His first score would place him at the 26-percentile of the Western Normal group and the second at the 74-percentile—a difference of 48-percentile ranks! Table IV shows the changes in the scores of the highest and lowest quartiles of this group.

TABLE IV. CHANGES IN SCORES OF SUBJECTS IN TOP AND BOTTOM QUARTILES

Subjects	Alpha Form		
	9	6	8
Top Quartile			
A	193	184	189
B	176	180	181
C	165	184	192
D	153	165	190
E	146	167	181
F	146	146	166
Mean	163.1	171.0	183.1
Percentile	89	94	98
Bottom Quartile			
S	103	139	161
T	97	126	146
U	97	124	129
V	94	129	136
W	89	142	153
X	88	149	143
Mean	94.6	134.8	144.6
Percentile	5	50	65

From Table IV it will be noted that the low fourth of these students make enormous gains from "practice"—about six times as much on the percentile basis as the top fourth. All of the subjects show significant gains as between the first and third tests. It is worth while to observe also that persons whose scores on

the first test are approximately or exactly the same show quite striking differences in later tests.

These results have bearing upon the consistency of the Alpha. Our results support and supplement those of Dunlap and Snyder. The importance of knowing beforehand how much previous test experience a subject has had before undertaking a treatment of his test record is emphasized. It may be shown later that true measures of intelligence can only be gotten in case the subjects have had sufficient practice with pencil and paper tests to take them approximately to their psychological limit of improvement through such practice. At any rate, before a comparison of the intelligence status, as represented by test scores, of any two or more groups is undertaken this factor should be investigated and suitable adjustment or allowances made therefore. It is not unlikely that many of the "sectional" differences between samples of like age and grade groups, between "superior" and "inferior" schools, and the like, are due at least in part to this particular skill in test-taking.

BIBLIOGRAPHY

1. YOAKUM, C. S., and YERKES, R. M. ed. *Army mental tests.* New York: Henry Holt and Company, 1920. pp. 161-83.
2. THURSTONE, L. L. "A cycle-omnibus test for college students," *Journal of Educational Research*, 4:265-78, November, 1921.
3. WOODWORTH, R. S. *Psychology. A study of mental life.* New York: Henry Holt and Company, 1921. 580 pp.
4. ARPS, G. F., and NOBLE, C. E. "University students' intelligence ratings according to the Army Alpha tests," *School and Society*, 11:233-37, February 21, 1920.
5. TERMAN, L. M. "Mental growth and the I.Q." *Journal of Educational Psychology*, 12:401-7, October, 1921.
6. WHIPPLE, G. M., and COBB, MARGARET. *Bulletin No. 17 of the Bureau of Educational Reference and Research, University of Michigan*, February 7, 1921.
7. DUNLAP, KNIGHT, and SNYDER, AGNES. "Practice effects in intelligence tests," *Journal of Experimental Psychology*, 3:396-403, October, 1920.
8. BROWN, WILLIAM, and THOMPSON, G. H. *Essentials of mental measurement.* London: Cambridge University Press, 1921. 216 pp.
9. PRESSEY, S. L. "Suggestions looking toward a fundamental revision of current statistical procedure, as applied to tests." *Psychological Review*, 27:466-72, November, 1920.
10. BURGESS, MAY AYRES, *The measurement of silent reading.* New York: Russell Sage Foundation, 1921. pp. 128-32.
11. KELLY, T. L. "The reliability of test scores." *Journal of Educational Research*, IV:370-79, May, 1921.
12. RUML, BEARDSLEY. *The reliability of mental tests in the division of an academic group,* (Psychological Review Monographs XXIV, No. 4.) 1917. 63 pp.

STUDY OF DEPTH AND RATE OF COMPREHENSION IN READING BY MEANS OF A PRACTICE EXPERIMENT[1]

ARTHUR I. GATES
Teachers College, Columbia University

In 1913 in discussing the frequency of improvement as the result of practice, Thorndike wrote: "So far as I am aware of the facts, no mental function has been deliberately practiced with an eye to improving it and with proper opportunity for the law of effect to operate without some improvement as a result".[2]

There are few if any reasons for supposing that the functions with which educational and mental tests are concerned are not similarly subject to the effects of practice. Since improvement is generally specific, the effects of practice should be discovered, lest quite erroneous notions of general growth be affirmed from what is merely specific improvement in the test itself.

With pupils of grades III to VI, inclusive, of the Scarborough School as subjects, the Burgess and the Thorndike-McCall reading tests, both of which have several supposedly equivalent forms, were used as materials for a practice experiment under conditions similar to those which ordinarily obtain in investigations of this type. The experiment with the Burgess test ran five days, five minutes per day; with the Thorndike-McCall thirty minutes per day for seven consecutive school days. Incidentally the data afford the possibility of computing the reliability and validity of the two instruments, as well as the relative difficulty of the several forms of each test.

THE VALIDITY OF THE TESTS

It will be advisable, first of all, to consider the validity of the two tests. What functions do they measure? Combining the

[1] A contribution of the Department of Educational Research of the Scarborough School, Scarborough, New York.

[2] Thorndike, E. L. *Educational Psychology.* New York: Teachers College, Columbia University. Vol. II. p. 151. 1913.

results of five trials of the Burgess and of seven Thorndike-McCall tests, measures of high reliability (consistency) are secured. The correlation between these two composite scores are as follows: grade III, 0.90; grade IV, 0.52; grade V, 0.48; and grade VI, 0.54. In grade III, the two tests appear to measure very nearly the same abilities, but such is not the case in other grades, where a valid distinction between the two functions seems to exist.

A considerable amount of individual work with beginners in reading, with exceedingly slow readers of high general intelligence according to the Stanford-Binet, and with readers of fair speed but less than average intelligence, has provided substantial evidence that (except in grade III or below) the Thorndike-McCall test measures *depth or power* of comprehension, in which rate is a minor factor, whereas the Burgess test measures *rate* of comprehension primarily.

Table I gives samples of situations frequently found among bright pupils whose reading is very slow. If a pupil has a high degree of general mental ability he will earn a high score on the Thorndike-McCall even if his reading is mechanically very poor; but even with great power or depth of comprehension his score on the Burgess will be low.

TABLE I. SAMPLE SCORES OF BRIGHT PUPILS WHO ARE SLOW READERS

Subject	Stanford I. Q.	School Grade	Grade Status Thorndike-McCall	Grade Status Burgess
A	125	VIII	9.2	3.0
B	122	VIII	11.2	3.2
C	119	V	5.1	2.7
D	109	VI	6.8	3.2
E	105	VII	7.4	3.1

We have found that intensive training of very slow readers produces marked improvement in scores on the Burgess test without perceptible improvement in the Thorndike-McCall (except in grade III or lower).

In grades below the fourth the two tests yield essentially the same results for many pupils because the Burgess material becomes a real test of depth of comprehension and because the Thorndike-McCall scores may be limited by speed and ability in word recognition. Effective training in the mechanics of reading—word recognition, etc.—sometimes results in an improvement in performances in both tests.

For purposes of further analyzing the group associations of reading and other abilities, the pupils of grades IV, V, and VI were combined, grade III being omitted for reasons just indicated. The correlations are shown in Table II.

The correlation between Burgess and Thorndike-McCall is 0.71, but when differences in chronological age are eliminated by partial correlation, it becames 0.59. This is a substantial correlation, indicating a fairly close association of *rate* and *depth* of comprehension in general. This fact in no way contradicts the findings presented above. The two tests measure essentially different even if correlated abilities; in other words, individuals high in one and low in the other are found but they are few. A coefficient of 0.59 allows for a great deal of irregularity.[3]

TABLE II. CORRELATIONS. GRADES IV, V, VI, COMBINED. ($n = 58$)

(a) Simple Correlations

	Burgess	Thorndike McCall	Stanford M. A.	Life Age
1. Burgess		0.71	0.56	0.40
2. Thorndike-McCall	0.71		0.73	0.58
3. Stanford M. A.	0.56	0.73		0.67
4. Life Age	0.40	0.58	0.67	

(b) Partial Correlations, First Order

	r
Burgess with Thorndike-McCall (Age eliminated)	0.59
Burgess with M. A. (Age eliminated)	0.41
Thorndike-McCall with M. A. (Age eliminated)	0.57

[3] The correlation 0.71 is higher than those reported for the grades separately for the reason that the range of ability in the single grade is much less wide.

As shown in Table II, the correlation of the Thorndike-McCall with the Stanford-Binet mental age is practically the same as that between the two reading tests. The association of Burgess with mental age is considerably less. The import of the correlation of Thorndike-McCall with mental age will be considered later.

The Influence of Practice Upon Rate of Reading as Measured by the Burgess Test

In order that the effects of practice might not be obscured by inequality in the difficulty of the several forms of the tests (although they are supposed to be of equal difficulty), a revolving schedule of practice was arranged. Each grade was broken into four groups, with the forms and practice distributed as follows:

Group	Forms Used on Indicated Days				
	First day	Second day	Third day	Fourth day	Fifth day
I	1	2	3	4	1
II	2	3	4	1	2
III	3	4	1	2	3
IV	4	1	2	3	4

Differences in difficulty of the forms were thereby neutralized because all four forms were used each day. The final average for each grade is an average of the means for the separate groups, thus giving equal weight to each form and each group.

Table III gives the results of five practice periods of five minutes each on successive days. The improvement, resulting from about twenty minutes practice (from the middle of the first period to the middle of the fifth) averages about fifteen percent. The figure is perhaps a trifle smaller than would be obtained from certain other functions such as arithmetic or cancellation, but it is nevertheless considerable. The improvement is specific; quite certainly rate of comprehension in general has not been improved to an equal degree. For that reason, certain cautions should attend the use of such a test as a periodic measure

of improvement in reading. Grade and age norms as published are doubtless based on results of *initial* tests. Repeated use of the Burgess test may well give a quite fictitious notion of general progress in terms of such norms. Although it may be that tests given at intervals of a month or semester will not produce improvement as great as when given on successive days, some discount for specific adjustment must be made. Among the entirely

TABLE III. PRACTICE EFFECT. BURGESS TEST

	First day	Second day	Third day	Fourth day	Fifth day
GRADE III					
Mean Score	4.1	4.0	4.2	4.4	4.5
Grade Equivalent	3.05	3.0	3.1	3.20	3.25
Percent	100	98	102	107	110
GRADE IV					
Mean Score	7.3	7.4	7.8	8.3	8.9
Grade Equivalent	4.80	4.9	5.3	5.8	6.4
Percent	100	101	107	114	122
GRADE V					
Mean Score	9.2	9.5	9.2	9.3	10.5
Grade Equivalent	6.7	7.0	6.7	6.8	8.0
Percent	100	103	100	101	114
GRADE VI					
Mean Score	13.2	13.3	13.2	14.6	15.2
Grade Equivalent	10.7	10.8	10.7	12.1	12.7
Percent	100	101	100	111	115
MEAN SCORE	8.5	8.6	8.6	9.2	9.8
PERCENT	100	101	101	108	115

uninitiated, improvement from practice will almost certainly be greater than that here disclosed for the reason that our pupils have had an unusually large amount of practice in various tests, including previous trials with forms one and two of the Burgess Scale.

PRACTICE EFFECTS IN POWER OF COMPREHENSION AS MEASURED BY THE THORNDIKE-MCCALL TEST

Four forms of the Thorndike-McCall test were distributed among four groups of pupils in each grade. The arrangement was similar to the one illustrated above for the Burgess test,

except that seven practice days, thirty minutes per day, were provided. Since, however, but four forms were used each day, the differences in difficulty of material were not altogether eliminated.

Table IV shows that seven days of practice produced no measurable improvement, a fact in conspicuous contrast to results ordinarily obtained in such experiments. The variability of performance from day to day was determined primarily by differences in the difficulty of forms (forms 7 and 8 being specially difficult); and when allowances for these differences are made, the practice curve for seven 30-minute periods is nearly flat.

The implication of this result is that the power of comprehension of printed words, among those mechanically able to read, is constantly practiced to its maximum; a maximum determined by the general mental maturity at the time. It may be that practice sufficient to develop maximal power of comprehension is supplied by the ordinary experiences of home and school. Even intensive specific practice produces no further increase in this ability; further progress depends entirely upon growth.

TABLE IV. PRACTICE EFFECT, THORNDIKE-MC CALL TEST

	First day	Second day	Third day	Fourth day	Fifth day	Sixth day	Seventh day
GRADE III							
Mean Score	35.2	36.7	36.8	33.6	33.3	34.3	34.5
Grade Equivalent	3.70	3.95	3.97	3.43	3.40	3.55	3.56
GRADE IV							
Mean Score	42.1	42.7	43.2	41.0	39.9	40.3	41.6
Grade Equivalent	4.85	4.95	5.03	4.67	4.46	4.55	4.77
GRADE V							
Mean Score	51.8	52.6	50.5	48.5	48.6	49.0	50.6
Grade Equivalent	6.47	6.65	6.25	5.92	5.93	6.00	6.27
GRADE VI							
Mean Score	58.3	58.7	57.9	56.4	57.5	57.9	59.1
Grade Equivalent	8.08	8.18	7.98	7.60	7.88	7.98	8.28
GRADES III-VI							
Mean Score	46.9	47.7	47.1	44.9	44.8	45.4	46.5
Grade Equivalent	5.65	5.78	5.68	5.32	5.30	5.40	5.58

Increase in the power of comprehension may depend upon the growth of native ability, which like height may be conceived to be determined by inner development independent of environmental factors save a sufficiency of food, exercise—ordinary healthful living—or, it may depend upon knowledge, breadth of information, and experience. In either case specific practice in the tests would be expected to produce little or no improvement.

Either conclusion, however, may be an incorrect explanation of the results. It may be that the units of measurement are so coarse that real improvement was made but not measured. Each unit in the test is equal to an age difference of from three to six months (the units among the highest scores are the equivalents of nearly two years). While it might seem quite reasonable to expect that three and one-half hours of specific practice would not equal three months or more of general experience, the fact is that even smaller amounts of specific practice in other functions do so, as we have already observed in the case of the Burgess test. We need, however, more intensive study with more minute scale units.

The correlation between scores on the Stanford-Binet and the Thorndike-McCall was found to be high. Since the Thorndike-McCall has carefully constructed Reading Age norms,[4] it will be instructive to compare the reading quotients (reading age divided by chronological age) with the intelligence quotients (mental age divided by chronological age). Among our pupils, the R.Q.'s give a higher median (128 compared with 122), a wider interquartile range (17 compared with 14), and a wider total range (72 compared with 50). The R.Q.'s depart widely from the I.Q.'s in particular cases, being higher in 36 cases, lower in 19, and the same in one. It is quite clear that despite the substantial correlation between these abilities, the quotients may not be used interchangeably. It is also true, although statements have been made to the contrary, that reading quotients very much higher than intelligence quotients may

[4] McCall, W. A. *How to measure in education.* New York: MacMillan Company, 1922. p. 73.

be obtained; in fact, among our pupils the R.Q. is higher on the average in two out of three cases.

The Reliability of the Tests

The Thorndike-McCall Test.—The criterion is the sum of the scores from seven forms of the test. The correlations based upon 64 pupils of grades IV, V, and VI between one or more tests and this composite are as follows: for one test, 0.84; for two tests, 0.90; for three, 0.93; for four, 0.95; for five, 0.96; and for six, 0.98. The correlations are sufficiently high to indicate very great stability of relative position among pupils when measured by different forms, especially if two or more forms are used. These data are, however, likely to be misleading if one is inclined to conclude that the uniformity of association in a group embracing three grades implies a security for the common practice of transmuting test scores into grade norms for the purposes of measuring individual or even grade progress during short periods.

By referring to Table IV and noting for each grade the highest and lowest grade equivalents, the reader will find that even when the large inequalities of the various forms are considerably reduced by using four different forms in each test, each grade as a whole shows a range of a half year or more between the extremes of the several tests. Variations from test to test will be larger than those indicated if a single different form is used at each trial. Since the various forms were distributed among the practice periods in such a way that practice effects would be fairly well equalized, even if they existed (which they apparently did not), the data of Table V show discrepancies which will occur in repeated tests due to inequality of the forms. If seven different forms are given on successive days and the scores transmuted into grade status, the average results for grade III vary over a range equivalent of 0.71 of a school year; for grade IV, 0.66 of a school year; and for grades V and VI, 0.86 and 2.5 years respectively. Forms seven and eight are especially difficult; Forms 2, 3, 5, and 6 are pretty uniform

in difficulty. While the tests have been carefully constructed and equated, the coarseness of the units gives even slight differences in difficulty a greatly magnified importance when transmuted into grade levels.

In order to determine the variability of performance of single individuals, with inequalities of forms fairly well eliminated, the data were assembled as follows: For each individual

TABLE V. SCORES AND GRADE EQUIVALENTS FOR DIFFERENT FORMS OF THE THORNDIKE-MCCALL TEST

	Form 1	Form 2	Form 3	Form 5	Form 6	Form 7	Form 8	Highest minus lowest
GRADE III								
Mean Score	35.0	36.9	35.7	36.6	36.7	32.1	31.8	
Grade Equivalent	3.65	3.95	3.84	3.93	3.95	3.3	3.25	0.71
GRADE IV								
Mean Score	42.3	42.6	41.6	43.0	42.6	41.2	38.6	
Grade Equivalent	4.88	4.93	4.77	5.0	4.93	4.71	4.33	0.66
GRADE V								
Mean Score	50.0	52.9	52.5	51.3	50.7	48.4	47.9	
Grade Equivalent	6.2	6.7	6.6	6.4	6.3	5.92	5.84	0.86
GRADE VI								
Mean Score	59.1	59.1	61.7	57.4	58.7	56.4	54.6	
Grade Equivalent	8.25	8.25	9.7	7.75	8.2	7.7	7.2	2.5
MEAN SCORE	46.6	47.9	47.9	47.1	47.2	44.5	43.2	

was computed the average score for the seven tests, and with this the score from the first test, the average of the first two tests, the average of the first three, etc. were compared. The deviations from the average of seven tests thus secured were averaged by grades. The results are shown in Table VI. Assuming that the average of seven tests is a perfect measure, the average individual deviates from it by 0.66 of a school year when he takes but one test, 0.41 years when he takes two tests, 0.38 for three tests, 0.32 for four, 0.24 for five, and 0.16 for six. The practical worker must consider the number of tests needed for his purpose. To measure the annual improvement of an individual one or two tests will suffice but to measure monthly improvement six tests at each examination are scarcely suffi-

TABLE VI. THE AVERAGE DEVIATIONS FROM THE COMPOSITE SCORE IF THE FIRST TEST, THE FIRST TWO TESTS, THE FIRST THREE TESTS, ETC. ARE CONSIDERED. THORNDIKE-MCCALL TEST

	Number of Tests					
	1	2	3	4	5	6
GRADE IV						
Scale Units	3.5	1.8	1.4	1.5	1.1	0.8
School Year	0.6	0.3	0.23	0.25	0.18	0.13
GRADE V						
Scale Units	3.1	2.2	2.2	1.4	0.9	0.7
School Year	0.6	0.44	0.44	0.28	0.18	0.14
GRADE VI						
Scale Units	3.2	2.0	1.9	1.7	1.5	0.8
School Year	0.8	0.5	0.48	0.42	0.37	0.2
MEAN						
Scale Units	3.3	2.0	1.8	1.5	1.2	0.8
School Year	0.66	0.41	0.38	0.32	0.24	0.16

TABLE VII. SCORES ON DIFFERENT FORMS OF THE BURGESS TEST, PRACTICE EFFECTS ELIMINATED

	Form 1	Form 2	Form 3	Form 4
GRADE III				
Mean Score	3.1	5.4	3.4	5.0
Grade Equivalent	2.6	3.7	2.8	3.5
GRADE IV				
Mean Score	7.9	9.6	6.5	7.5
Grade Equivalent	5.4	7.1	4.3	5.0
GRADE V				
Mean Score	9.5	12.3	8.2	8.4
Grade Equivalent	7.0	9.8	5.7	5.9
GRADE VI				
Mean Score	14.1	15.8	13.1	12.5
Grade Equivalent	11.6	12.+	10.6	10.0
GRADES III-VI				
Mean Score	8.7	10.8	7.8	8.4
Grade Equivalent	6.2	8.3	5.3	5.9

cient. If his purpose is to measure the progress of a grade as a whole, fewer tests are required provided several forms are used at each test. A good method is to employ the same forms at each test. For example, if three tests are to be made annually, give each third of a class a different form at the first test, use the same forms but give them to a different third of the class on the second and third tests, thereby eliminating differences in difficulty among the forms.

The reliability of the Burgess Test.—Since all four forms of the Burgess test were used in each of the five trials, the influence of practice is equal for each form. Table VII shows the relative difficulty of the forms in terms of the number of paragraphs read and in terms of the grade equivalents. The same pupils secure on Form 2 a score equivalent to three grades higher than that earned on Form 3. This means that semi-annual or even annual progress cannot be adequately measured by using different forms at each test. In ordinary work, the effects of specific practice on the tests, would be added to the discrepancy here shown.

The variability of the performances of an individual on this test is discouragingly great. Table VIII gives as samples the scores obtained by the five pupils first in the list of Grade IV. Subject 1, reads 10 paragraphs in one test and 4 in the next; subject 2, ranges from 7 to 10; subject 3, from 3 to 6; and subject 4, from 10 to 17.

From such data as shown in Table VIII, the average score for each individual and the average deviation of his scores

TABLE VIII. SCORES OF THE FIRST FIVE PUPILS OF GRADE IV

Day	Form	Scores of first five pupils				
		1	2	3	4	5
1	1	6	8	5	11	5
2	2	10	10	5	12	6
3	3	4	7	3	10	5
4	4	12	7	5	15	6
5		9	8	6	17	5
Mean Score..		8	8	5	13	5
A. D. from Mean.....		2.6	0.8	0.8	2.4	0.4

from his mean were computed, as shown in the last two lines. Averaging these two lines for each grade, we have Table IX. If we give the typical individual of grade IV, V or VI five tests on succeeding days, half of the single scores will differ from the average of all scores by an amount equivalent to a grade and a half or more. For grade III the amount is less, half a grade. Measuring monthly or semi-annual progress of an individual by a single test is out of the question.

The only remedy, taking the Burgess tests as they are, is to use two or more forms. In Table X are shown the average deviations from the mean scores of five tests when a single test,

TABLE IX.—THE AVERAGE DEVIATION OF THE TYPICAL PUPIL'S SCORES FROM THE MEAN SCORE OF HIS FIVE TESTS. BURGESS TEST

	Grade			
	III	IV	V	VI
Mean Score	4.3	7.9	9.7	13.9
A. D. from Mean	1.0	1.2	1.8	1.5
Grade Equivalent of A. D.	0.5	1.2	1.8	1.5

TABLE X. THE AVERAGE DEVIATIONS FROM THE COMPOSITE SCORE (FIVE TESTS). BURGESS TEST

	Mean Score 5 Tests	Average Deviations for:			
		1 Test	2 Tests	3 Tests	4 Tests
Grade III					
No. Paragraphs	4.3	1.0	0.6	0.3	0.2
School Years		0.5	0.3	0.15	0.1
Grade IV					
No. Paragraphs	7.9	1.2	0.8	0.7	0.5
School Years		1.2	0.8	0.7	0.5
Grade V					
No. Paragraphs	9.7	1.8	1.7	0.6	0.3
School Years		1.8	1.7	0.6	0.3
Grade VI					
No. Paragraphs	13.9	1.5	1.1	0.7	0.3
School Years		1.5	1.1	0.7	0.3
Mean Number					
Paragraphs		1.4	1.05	0.57	0.32
School Years		1.25	0.97	0.54	0.3

the average of two tests, etc. are obtained. For a single test the deviation is 1.25 school year; for two tests, 0.97; for three 0.54; and for four 0.3. The average of all four tests is no more reliable than measures of semi-annual progress demand.

Because individuals differ so widely in reading ability the relative positions are fairly maintained in different tests by the Burgess, even if the gross scores for each vary so much as to make it difficult to locate a pupil's grade status. Using the sum of the scores for five tests as the criterion the correlation of one test with the criterion is 0.86; of two tests, 0.92; of three tests, 0.95; and of four tests, 0.96.

THE INDIVIDUAL LABORATORY VERSUS THE DEMONSTRATION METHOD OF TEACHING PHYSICS

E. W. KIEBLER
Hillsdale High School
AND
CLIFFORD WOODY
University of Michigan

The purpose of this experiment is to set forth the results of an attempt in the Hillsdale high school (Michigan) to evaluate the merits of the individual laboratory method and the demonstration method of teaching physics. In the individual laboratory method each student or group of two students performs each experiment while in the demonstration method the instructor, with the aid of one or two students, demonstrates the experiment and the students get its import through observation and discussion rather than through actual performance. No difference is manifest between the two methods in the early part of the experiment. With both methods the same preliminary reading assignments are made and when the classes assemble the same type of short quizzes are given to test the effectiveness of their reading and their understanding of the new terms of the experiment. It is from this point on that the procedure with the two methods is different.

With the individual laboratory method, each student begins to study the experiment as described in the laboratory manual and prepares a "temporary sheet" embodying a definite statement of the object of the experiment, a list of the apparatus to be used, and a table suitable for recording the data to be obtained. After the "temporary sheets" have been prepared, a quizz is given on "methods of performance" and special emphasis is placed on necessary precautions and possible sources of error. The students now pass to the laboratory and individually or in pairs perform the experiment, record the necessary data, make the required mathematical calculations, and write up the "temporary discussions." Later, after the "temporary sheets" have been corrected

and the general errors have been discussed in class, each student writes up the experiment in permanent form.

With the demonstration method, after the short quizz on the assigned readings and new terms, the instructor announces the experiment for the day. He then asks each student to write on the "temporary sheet" the purpose of the experiment. He questions the students concerning the apparatus needed and each lists the necessary pieces. Through more questions he develops the method of procedure and emphasizes necessary precautions and possible sources of error. Then in the presence of the class and with the assistance of one or two members he performs the experiment. Each student copies the data which he writes on the board, makes the mathematical calculations from them, and completes the required "temporary discussions." Later as with the laboratory group, when the "temporary sheets" have been corrected and the errors discussed, each student writes up the experiment permanently.

Thus with the two methods the same assignments are made in the textbook, the same references are consulted, the same discussions precede and follow the experiments, and the same "temporary" and "permanent" write-ups are demanded. The main difference between the methods is that one group prepares the "temporary sheets" from the laboratory manual and individually performs the experiments while the other group develops the methods of procedure in class without a laboratory manual and observes the performance of the experiments by the instructor and one or two students.

In attempting to evaluate these two methods of teaching physics, the subject matter embodied in fourteen experiments on the topic of "heat" was taught to one group of students by the individual laboratory method and to the other group by the demonstration method. The two groups consisted of most of the students regularly taking physics and were equated on the basis of scores on the Army Alpha Test. The average score for one group was 122.82 points on the Army Alpha Test and for the other group, 122.94 points. Each group contained good, average,

and poor students, but only those whose attendance was perfect and whose attitudes were conducive to good work.

The fourteen experiments emphasized in this investigation were as follows:

1. To determine a source of heat.
2. The liquid and air thermometers.
3. To test the fixed points on a centigrade thermometer.
4. To determine the coefficient of linear expansion.
5. To determine the specific heat of a metal.
6. To construct a curve of cooling through change of state.
7. To determine the heat of fusion on ice.
8. To determine the effect of dissolved solids on the temperature of liquids.
9. To determine the heat of vaporization of water.
10. To ascertain dew point.
11. To compare the conductivity of metals.
12. To show the poor conductivity of liquids.
13. The laws of radiation.
14. To determine the mechanical equivalent of heat.

With the first seven of the above experiments, the individual laboratory method was used with one group "X" and the demonstration method with the other group "Y," but with the last seven experiments the methods were reversed.

Three different types of tests were given in the attempt to measure the effectiveness of the two methods: (1) tests for temporary or immediate knowledge of the facts of the experiments; (2) tests for permanent or delayed knowledge; and (3) a test for knowledge of how to apply the principles and technique learned in attacking a new problem. In testing for temporary or immediate knowledge, an informal test was given to each group from one to two weeks following each experiment. In testing for permanent or delayed knowledge a general test over the contents of the first seven experiments was given two weeks after the completion of the seventh experiment. A similar test was given over the contents of the last seven experiments two weeks after the completion of the fourteenth experiment. In the case of two of the experiments, as a variation to this method of testing for immediate and delayed knowledge, the tests for immediate knowledge were given upon the completion of the "tempo-

rary discussions" and were given again two weeks later after the experiments had been permanently written up.

In the test for knowledge of how to apply the principles and technique learned, the students were confronted with a new problem, viz., "To find the temperature of a white hot body." They were furnished with a list of the apparatus needed for the experiment and were then given an informal test in which they were asked to make a simple drawing showing how the apparatus would be used and to fill out broken sentences describing a possible method of procedure. Each student in the individual laboratory group then performed the experiment while in the other group the instructor gave a demonstration. Two weeks later both groups were rigorously tested again.

All tests used in attempting to evaluate the two methods might be termed *informal* tests. They were, however, as nearly objective as possible, and they embodied many of the principles utilized in the construction of standardized tests. Much use was made of the "completion test," the "matching test," the "true and false test," and the "best answer test." A few abbreviated samples will make clear the nature of the tests and will indicate the accuracy and the thoroughness with which the work was appraised.

SAMPLE I

The words at the left are closely related to some expression at the right. In the parentheses before each word place the number of the expression which you think is most closely related.

( ) Expansion	1. One one-hundredth the distance between the freezing and boiling points on a centigrade thermometer
( ) Error	2. The number of units of heat required to raise the temperature of one unit mass one degree
( ) Degree	3. Rise of mercury in a thermometer
( ) Specific Heat	4. Loss of heat by copper beaker to the air
( ) Radiation	5. The difference between the accepted value and the obtained value

Sample II

Complete the following statements:

1. ———————— is a form of molecular energy.
2. The paraffin melted at ———— points because the wires differ in ————————

Sample III

Some of the statements concerning humidity are true and some are false. If the statement is true, draw a line under TRUE. If it is false, draw a line under FALSE.

Warm air always contains more moisture than cold air. TRUE FALSE

Sample IV

Place a check mark ($\surd$) in the parentheses at the left of the statement that best completes the sentence.

Tin-plate tea-kettles often have copper bottoms because

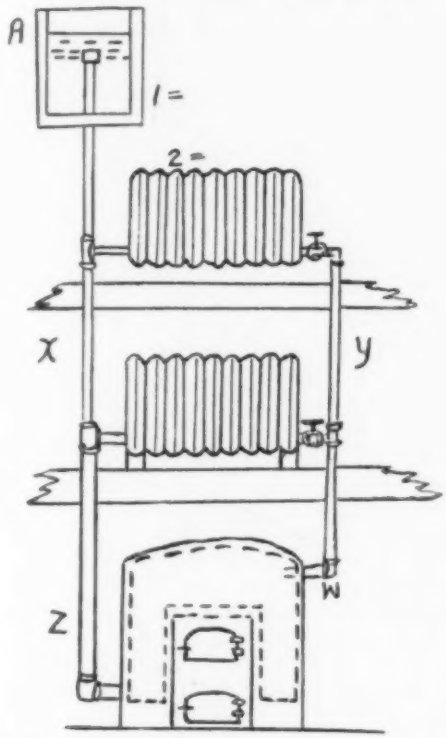
( ) They will not rust so easily.
( ) Tin plate does not burn out or rust out so easily.
( ) Copper being a soft metal is more easily joined into a tight seam around the bottom edge.
( ) Copper is a better conductor of heat.

Sample V

Complete the following data:

Temperature of the room	19.0 degrees centigrade
Mass of calorimeter	75.5 grams
Mass of water and calorimeter	218.0 grams
Initial temperature of water	6.0 degrees centigrade
Final temperature of water	33.0 degrees centigrade
Mass of water, calorimeter, and steam	225.0 grams
Initial temperature of steam	100.0 degrees centigrade
Heat of vaporization	—— calories
Accepted value	—— calories
Error	——

SAMPLE VI



The drawing at the left represents a hot water heating plant in the winter with no fire in the boiler. The level of the water in the plant is represented by the line "A" under these conditions. Draw a line parallel to "A" to indicate the level of the water in the plant in July (no fire) provided no water has been added and none removed. Label this line "B." Indicate by another line parallel to "A" the level of the water in winter with a hot fire in the boiler and with no water having been removed or added. Label this line "C."

Opposite the number "1" write the name of this part of the plant. Opposite the number "2" write the name of this part of the plant.

At the points "x," "y," "z," and "w" indicate by arrows the directions in which the water would circulate.

Table I, exhibiting the results of the tests for immediate knowledge in the fourteen experiments, shows that the mean score with the demonstration method was 60.52 percent correct; with the individual laboratory method, 59.66 percent correct. In the first seven experiments the group with the demonstration method was superior in five experiments and inferior in two experiments. In the second seven experiments, the group with the demonstration method was superior in three and inferior in three experiments while in one experiment the scores for the two methods were the same. It is interesting to note that those using the

TABLE I. MEAN PERCENTS ON TESTS FOR IMMEDIATE KNOWLEDGE

Number of Experiment	Demonstration Group "x"	Laboratory Group "y"	Number of Experiment	Demonstration Group "y"	Laboratory Group "x"
1	82.6	74.01	8	76.81	75.00
2	75.8	70.00	9	56.80	66.56
3	51.8	44.17	10	36.02	40.06
4	44.14	53.00	11	79.28	79.28
5	53.17	45.60	12	59.50	61.10
6	43.84	43.27	13	71.49	70.73
7	43.39	44.10	14	72.73	68.43
Mean.....	56.39	53.49	Mean.....	64.66	65.87

demonstration method during the first seven experiments were more successful with the laboratory method during the last seven experiments. The apparent superiority of this group cannot be charged to superiority in intelligence for, according to the intelligence test, the two groups were equal. This at least shows that the demonstration method has not been a handicap to those students. It should be added that the individual laboratory method is superior in those experiments which are difficult to demonstrate. Experiment No. 14 on "Determining the Mechanical Equivalent of Heat" proved to be very difficult for the students, but the procedure was easily seen by them and consequently the demonstration method gave superior results. In experiment No. 4 on "Determining the Coefficient of Linear Expansion" it was almost impossible for the class to observe the micrometer and consequently the individual laboratory method gave superior results.

TABLE II. MEAN SCORES FOR TESTS FOR PERMANENT KNOWLEDGE GIVEN

	Demonstration	Laboratory
Test two weeks after completing first seven experiments.....	60.3	61.6
Test two weeks after completing second seven experiments.....	56.8	57.0
Mean for both tests	58.51	59.3

Table II, showing the mean scores on all tests for permanent knowledge, shows a slight advantage for the individual laboratory method, although the advantage is so slight that it is of little consequence.

Table III, which compares the mean scores made with the two methods on the test given immediately after the temporary sheet had been completed and repeated two weeks after the experiment had been written up, shows higher scores for the demonstration method in three of four possible comparisons. The higher scores in the first tests with the demonstration method may mean that the facts of the experiment are more vividly portrayed than with the individual laboratory method. The fact that the group using the individual laboratory method made much

TABLE III. MEAN SCORES OF TESTS FOR IMMEDIATE KNOWLEDGE GIVEN UPON COMPLETION OF TEMPORARY SHEET AND OF TEST FOR PERMANENT KNOWLEDGE GIVEN TWO WEEKS LATER AFTER THE EXPERIMENT HAD BEEN WRITTEN UP

Number of Experiment	Demonstration Scores		Laboratory Scores	
	First Test	Second Test	First Test	Second Test
7	40.2	43.4	36.0	44.1
14	62.7	72.7	48.0	68.4

more gain in the second test than did the group using the demonstration method is more evidence of the initial superiority of the demonstration method. The gains shown on the second test by both groups attest to the value of permanently writing up the experiments.

Table IV, showing the mean scores obtained by the two groups when confronted with the task of solving a new problem

TABLE IV. MEAN SCORES ON TEST FOR KNOWLEDGE OF HOW TO APPLY PRINCIPLES AND TECHNIQUE

	Demonstration Group "x"	Laboratory Group "y"
Test on method or procedure	57.86	51.7
Test after performance of experiment	77.58	75.4
Gains	19.72	23.7

when only the list of apparatus was given, indicates that the group taught by the demonstration method is superior on the test on method of procedure before the experiment was performed and on the test after the performance of the experiment. This fact seems to make clear that the group using the demonstration method has as much ability in attacking new problems as the group trained with the individual laboratory method.

On the basis of the foregoing data, the following observations seem justified:

1. The results secured through the use of the demonstration method were as good, if not superior, in all three types of tests given—the tests for immediate knowledge, for permanent knowledge, and for knowledge of how to apply principles and technique in solving a new problem. While the advantage in favor of the demonstration method was small, it was very significant.

2. The individual method tended to be superior in those experiments that are especially difficult to perform or in which great care must be exercised to see the exact procedure. Such facts suggest that the most effective method depends upon the nature of the experiments themselves and suggest the need for scientifically classifying them.

3. When the demonstration method gives equal or superior results, it is to be preferred to the individual laboratory method because it saves about one-half of the time usually devoted to performing the experiments and permits the instructor to use the time thus saved in relating the facts and principles to allied phenomena. Furthermore, the enthusiasm of the class working together is most valuable.

4. New problems involving principles and technique already learned arouse much interest and enthusiasm and offer opportunity for rich returns in teaching.

5. The results of the experiment suggest that, in the smaller high schools at least, better teaching would result and much money would be saved if the laboratories could be arranged and equipped so as to facilitate demonstration of many of the experiments instead of trying to provide apparatus for the individual performance of all of them.

Editorials

AN UNRECOGNIZED ABILITY IN TEACHING

What do you mean by a good teacher? If answers to this question could be obtained from all the readers of the JOURNAL OF EDUCATIONAL RESEARCH, their variety would no doubt be very great. And if to these replies were added the more formal answers of the score-card makers—Boyce, Elliott, Landsittel, Connor, Rugg, *et al.*—surely we might then expect to have in hand every item upon which an estimate of teaching ability could reasonably be based. We should certainly find frequent mention of *personality, scholarship, method, discipline, leadership,* and *cooperation.* Even *voice, good looks,* and *sense of humor* would be abundantly represented; while those who look to the pupil rather than to the teacher for the real measure of teaching would name a multitude of desirable knowledges, skills, activities, ideals, appreciations, and attitudes which good teaching should develop.

It is probable, however, that no one would include in his list an item having to do with the number of pupils a teacher can successfully instruct. We are content to list the qualities which condition good teaching without inquiring how many pupils can be effectively influenced by these qualities. We estimate the results of teaching without taking account of the number of pupils who exhibit these results. Is it too much to say that if one teacher can teach twenty-five pupils while another can teach fifty equally well, the latter is worth twice as much as the former? And if he is worth more, should he not be better paid?

That differences exist among teachers in respect to the number of pupils they can teach successfully is certain. We should expect this to be true from all we know of individual variation

in other respects. Moreover, these differences are matters of common observation.

We lately observed the teaching of a class of about seventy pupils. The exercise was definite, interest was strongly aroused, the attention of all the pupils was engaged, and their intellectual participation was evident. It was not a lecture, but a series of questions, answers, and discussions in which everybody took part, not merely when reciting but through active listening. Independently of his qualifications in other respects, the teacher's success in reaching every individual in this large group stamped him as exceptional.

On another occasion a teacher of a group of about the same size was found to be quite unable to reach beyond the middle of the room. Students on the back seats and on the wings of the class were passing notes to each other, studying the assignment for another hour, and otherwise following their own devices. It is submitted that whatever personal assets this teacher may have possessed—whatever of scholarship, method, or personality—he was not a successful teacher of a large group.

Probably the second teacher was not as good a teacher as the first in any respect; but curiously enough, it was of the "absurd" size of the class that he complained. On the other hand, the first teacher preferred a large class and declared that he could get little inspiration out of a small class. The only objection he had to a large class was the clerical work involved—the taking of attendance, the writing of scholarship reports, and the rating of papers. It would have paid the institution to hire that work done for him and to give him the large classes which he was capable of teaching and which he really enjoyed.

At the risk of digressing we wish to remark that we have never fully understood the attitude of educational administrators in insisting that the teacher do all the clerical duties connected with his instructional work. There seems to be a feeling that one cannot teach well unless one pores over test papers. Some attention to them is necessary, but much reliance can be placed on the proper reports made by readers of these papers. Does

the business executive fail accurately to administer his affairs because he does not read all his letters, handle all his orders, or study every inventory and payroll? He finds he can attend to larger and larger enterprises by relegating these activities to others who are paid to handle them and to keep him informed about them. So perhaps the teacher could handle larger and larger affairs—larger classes for one thing—if his energies were released for the real job of teaching.

But at present teachers are neither trained, selected, nor judged with reference to their ability to teach many rather than a few pupils.

In the normal school they are taught in the theory courses that small classes are to be desired at all events and at all hazards. In their practice teaching they are given small groups of children—frequently from ten to fifteen—and their efforts are adjusted to small-class situations.

As to the selection of teachers we doubt whether superintendents are really considering the ability of candidates to handle large classes. Yet each superintendent has but a limited amount of money which he may spend for teachers' salaries. If he can secure teachers who can handle large classes, he can either pay better salaries or have some money left for other things. In all our experience, however, with the methods of superintendents in employing teachers we do not recall an instance where a candidate was asked the size of the class he could teach or liked to teach. Nor do the blanks which are used in connection with appointments and on which information regarding the experience of teachers is recorded contain a single item designed to bring out the success of the candidate in handling large groups of children.

Finally, as we have already indicated, the same neglect of this rather obvious matter is evident in the schemes for rating teachers. We lately obtained information from a large number of superintendents as to their plans for appraising the work of teachers. In no single instance did they estimate the teacher's value to the school system on the basis of his success with many children.

The fact is that the educational world has been assuming that good teaching cannot be done in large classes. Superintendents of school systems, presidents of colleges and universities, and in fact, educators in general have employed much of their time in bemoaning the size of the classes under their jurisdiction. Two remedies are possible, and only one of them has been tried. The first is to reduce the size of classes. This has been everywhere attempted. The other is to secure large-class teachers. This has not been tried, and it should be. The assumption seems to have been made not merely that large classes are an evil, but that small classes—no matter how small—are a blessing to be sought at the sacrifice of almost anything else.

Seldom does the administrator have the unqualified choice between the large class and the small class. No factor enters so decisively into the cost of education as class size. Accordingly, the administrator must choose one or the other of the following situations: first, the presence of a large number of teachers each receiving a small salary and teaching a small class or second, a smaller number of teachers receiving larger salaries and teaching larger classes. In other words, the cost element has been left pretty largely out of account, while the theoretical aspects of class size have held the center of attention.

If superintendents and other administrators would demand teachers who could reach the individual student and secure his intellectual participation despite the fact that he is a member of a large class, and if they would make it known that large-class teachers would be paid generously, we believe that some financial difficulties would be diminished without necessary detriment to the quality of the instruction. At any rate, we feel sure that there are small-class teachers and large-class teachers, and that the value of the latter has been too little regarded.

B. R. B.

Reviews and Abstracts

PRESSEY, S. L. AND L. C. *Introduction to the use of standard tests.* New York: The World Book Company, 1922. 263 pp.

In many instances books dealing with standard measurements present the subject in such technical fashion that the beginner is likely to become hopelessly bewildered. The recent book, *Introduction to the Use of Standard Tests*, by Sidney L. and Luella C. Pressey, has gone to the other extreme.

In attempting to realize their purpose of presenting a simple manual for tests and showing their practical use, the authors have omitted or made only a brief reference to topics which seem essential to workers in this field. The reader is given the impression that the intelligence quotient of any child is the determining factor in his grading. Action based upon this assumption would lead the principal or superintendent into serious error. It is the educational age rather than the mental age which shows the grade to which the child belongs. The achievement quotient indicates the excellence of previous work, while the intelligence quotient shows capacity for future progress.

McCall says: "Educational age is superior to mental age for this purpose (classification) because it and it alone reveals directly what pupils are of equal status educationally. Mental age measures educational status only indirectly." The same point is made very emphatic by Burton in "Supervision and the Improvement of Teaching."

One of the stated purposes of the book is to assist teachers, untrained in the use of tests, to handle test results. How this can be done without consideration of mode and the normal curve of distribution is difficult to understand. Both of these topics are omitted. Correlation is treated in a most inadequate fashion. To say that "a coefficient of + 1 shows perfect relationship" and omit all reference to its derivation only confuses the beginner for whose help the book it written.

Those who undertake a testing program for the first time rarely concern themselves with the principles of test construction. They want to know how to tabulate and interpret test results. The authors devote Chapter XIII, sixteen pages, to the topic "How Tests are Made," and make no reference to the essential element of variability, save for a brief discussion of "scattering". This term is by no means common, and is only a step in explaining mode and variability. They fail to carry it far enough. Determination of the median does not give a complete picture of class conditions. Two classes might show identical central tendencies with the scores of one class ranging from second- to eighth-grade standards, while the second class may have practically no

variation from the median. Quartile deviation would seem almost essential to an understanding of the situation.

Any superintendent about to inaugurate a testing program will find Chapter XIV most suggestive. The authors have indicated the successive steps with such detail that failure would seem impossible. Equally valuable is the chapter dealing with the method of making testing programs worth while. In the judgment of the writer this would be much more beneficial had some consideration been given to the question of graphical representation. Results shown in this way are sufficiently impressive to justify thorough discussion.

Very wisely, space is economized by omitting all description of tests. Representative tests in each subject are named, leaving the reader to ascertain details by a study of the tests themselves.

The Pressey tests have proved so useful that users of these tests were possibly led to expect more of the book than was reasonable. The treatment of the subject is too inadequate and superficial to serve as a complete manual for the practical school man. Its chief value will be found in its use as a supplement to other and more comprehensive discussions.

D. C. Bliss

Superintendent of Schools, Montclair, N. J.

Counts, George Sylvester. *The selective character of American secondary education*, (Supplementary Educational Monographs, No. 19.). Chicago: University of Chicago Press, 1922. 162 pp.

In this book Doctor Counts has made a highly valuable addition to the knowledge that we have from the former studies by Vandenburg, Holley, Thorndike, Proctor, Book, Burdge, and others concerning pupils in our high schools. For 17,992 puipls in the high schools of Bridgeport, Mt. Vernon, St. Louis, and Seattle, and also for 514 children of high-school age at work in Seattle, he reports the parental occupation and its relation to enrollment, progress through school, the curriculum elected, and declared expectations following graduation. In separate chapters he gives additional information on the public high school and the cultural level, family influences, the immigrant, the negro, and psychological data, with similar facts for two private secondary institutions, the University of Chicago High School, and Phillips Exeter Academy. All these chapters are summarized and then followed by a cautious essay on the high school and democracy.

Undoubtedly the tremendous growth during the past generation of high-school enrollment has increased the heterogeneity of the pupil population and probably has also lowered the average of intellectual ability. But Counts shows that, at least in the schools studied, the children of families of superior wealth and social position enter high school in larger proportion and remain longer than do their less fortunate fellows. This is of course to be expected so long as curricula and courses of study are more suited to their capacities, aptitudes, and probable future needs, although this is far from explaining all of the increase in persistence from the grammar grades. Apparently the ideal forcibly presented by the Commission on the Reorganization of Secondary Education,

that the high school is for all children from twelve to eighteen years of age, has been more fully adopted by the pupils than by the receiving schools.

The table on page 33 is perhaps the best single summary of the monograph. It shows the number of children in the high schools of four cities from each of fourteen occupational groups for every 1000 males over 45 years of age engaged in that occupation in these cities. For 1000 parents in managerial service the number of pupils is 400; for those in professional service, 360. At the other end, for 1000 parents in personal service the number of pupils is 50, and for those in common labor, 17. The percent of children from the superior social and professional classes increases sharply from the freshman to the senior year, while that for children of parents in the trades, transportation, public and personal service, and in common labor declines. For every 1000 pupils in the freshman year whose parents are in professional service, 602 remain four years later; for every 1000 freshmen whose fathers are common laborers, 124 remain. The special study of 739 children in the sixth grade at Mt. Vernon shows that the selective action begins even before the high-school age.

It may be said that in a general way every observant secondary-school man already knew of this selective process. But it is a valuable service that Doctor Counts has rendered in presenting exact facts which must be more or less true for high schools the country over. Like Burdge's "Our Boys," this study by Doctor Counts is so filled with fundamental data basic to the direction of the new programs of secondary education that it is difficult to overemphasize its importance. It should receive close study and profound reflection by all who are concerned with promoting a sound adaptation of high schools to the increasingly heterogeneous body of pupils seeking higher education; and unless the principles of democracy are denied, its facts must lead to a modification of curricula, courses of study, methods of teaching, and social education.

The study is planned with great care, and its conclusions are drawn with scrupulous qualification. The only sections that may fairly be impeached in these matters are those on first- and last-born children, on negro pupils, and on ethnic groups. On the first topic Counts realizes at least a part of his difficulties, in the second he is more generous in his interpretation than the results of other studies would seem to warrant, and in the third he presents facts without a proper consideration of modifying factors. Commendation should be made of the unusually clear and helpful graphs supporting the text.

THOMAS H. BRIGGS

Teachers College, Columbia University

GRAY, WILLIAM SCOTT, and others. *Remedial cases in reading: their diagnosis and treatment*, (Supplementary Educational Monographs, No. 22.). Chicago: University of Chicago Press, 1922. 208 pp.

At a time when investigations in reading run over-much to a mere statistical treatment, this intensive study of individual children is particularly to be welcomed. Many of those who have been interested in research in reading have despaired somewhat of squeezing much more juice out of the compila-

tion of the results of tests and experiments administered rather crudely to masses of children. Under the influence of the praiseworthy desire to get an adequate sampling of cases, there has been a tendency to forget that large numbers of cases are significant only when the data for each and every case have been selected in such a fashion as to give a record of all the important relevant facts. This monograph is also valuable in turning the attention of teachers to the need of adapting instruction to the individual. In the enthusiasm for some of the improved methods of teaching silent reading there is a tendency to use these methods in a wholesale fashion without sufficient adaptation to the individual child. The data in this monograph are of special significance in pointing out the critical nature of the difficulties of exceedingly poor readers. At the same time, the remarkable results obtained by Dean Gray and his associates are a real inspiration to teachers to work harder and more intelligently with such children.

The monograph reports the detailed diagnosis and remedial treatment of each of 27 pupils at or near Chicago. It also reports a summary of the remedial work done by Dean Gray with the cooperation of Miss Wilson, assistant superintendent of schools at Toledo, Ohio. For each of 27 pupils there is reported: (1) a general description of the case; (2) a preliminary diagnosis consisting of an intelligence test and a battery of reading tests; (3) a detailed diagnosis including a vocabulary test and a detailed analysis of various types of oral and reading difficulties; (4) a study of eye-movements; (5) a description of the remedial instruction; (6) results of the remedial instruction, reported in tabular form, and an interpretation of these results.

The pupils are described in certain groups under the following heads: (1) pupils who have made little or no progress in learning to read; (2) pupils who encountered serious difficulties in interpretation; (3) pupils who encountered difficulties primarily in the mechanics of reading; (4) pupils who encountered difficulties in the rate of silent reading; (5) pupils who were weak in practically all phases of silent reading.

The remedial work used by Dean Gray is skillfully organized; and it is in harmony with the results of modern experiments and the best classroom practice. His constant emphasis upon attention to the thought and to interpretation is of especial significance. It is the foundation of his attack in the case of these deficient pupils, as it should be in the attack of those who are teaching ordinary class work in reading. The comprehensiveness of his remedial work can be illustrated by quoting his conclusions concerning remedial instruction for pupils who encounter serious difficulties in interpretation:

"An analysis of the reports of individual cases in this chapter shows clearly that there are certain suggestions which may be followed to advantage in organizing remedial instruction for any pupil who encounters difficulties in interpretation. (1) The interest of the pupil must be aroused and his cooperation secured. (2) Selections must be used which are directly related to the interests of the pupil. (3) A factual background must be developed in order to secure accurate interpretations. (4) Specific problems must be assigned during the early part of the training period in order to direct the pupil's

attention to significant points. (5) The problems which are assigned at first should be relatively simple and should increase in difficulty as rapidly as the pupil's power of interpretation develops. (6) Different types of problems must be assigned from time to time in order to insure growth in all important phases of interpretation. (7) Responsibility for independent interpretations must eventually be imposed on the child. This means that he must re-read or study passages several times, if necessary, in order to secure adequate interpretations. (8) The cooperation of the teachers of content subjects must be secured in order to insure the establishment of good habits of interpretation in all reading activities."

Every one of these remedial measures is valuable for the teaching of reading in ordinary normal classes. In fact, the summary of causes of failure for various classes of poor readers and the summary of remedial measures constitute an excellent outline of a sound course of study in reading.

This monograph is a valuable addition to the professional library of anyone who is concerned with the problem of teaching reading. It maintains the usual high standards of various Supplementary Educational Monographs which have appeared as a result of the research on reading at the University of Chicago. In addition to being a valuable scientific contribution, the sympathetic study and patient work done by Dean Gray in helping these children constitute a valuable model for emulation by the supervisor and classroom teacher.

It would be a great thing for education if all recommendations as to methods of teaching should be based, as is this, upon actual experimental teaching by the professor who advocates them.

ERNEST HORN

State University of Iowa

GRAY, CLARENCE TRUMAN. *Deficiencies in reading ability: their diagnosis and remedies.* New York: D. C. Heath and Company, 1922. 420 pp.

Most readers of the JOURNAL OF EDUCATIONAL RESEARCH can recall that among their elementary school classmates there were some pupils who were slow in reading. Those pupils were sometimes prodded, sometimes kept in school at recess, and sometimes only continued in their first readers. The reviewer remembers that a very successful country merchant received the last mentioned treatment. In four successive years he wore out four first readers!

The days referred to above are comparable to the days of turnkeys and bloodletting. No longer need slow readers go unaided. The modern teacher can equip herself for the intelligent diagnosis of pupils' reading ailments and for the prescription of suitable remedies.

In order to undertake such diagnosis of reading ailments, the teacher must be equipped with proper instruments. It happens that such instruments have been used more or less uncritically by some of us in our diagnostic work. It is, therefore, necessary that we understand the instruments which we are to use in diagnosis. To this end, Doctor Gray has provided for us an analysis of reading tests and of reading itself from the standpoint of these tests. Then follow analyses of reading from the standpoints of visual perception, the motor

processes of reading, and certain higher mental activities which accompany reading. These analyses enable the author to set up the only sound basis for directing or supervising the study of any school subject, i. e., a knowledge of processes involved in the pursuit of the subject.

The second division of the book contains a compilation of tests and methods of observation for diagnostic purposes. Here is found a group of standard oral and silent reading tests with suggestions for their use. The value of this portion of the book lies chiefly in the fact that it provides an array of instruments of diagnosis, together with a discussion of special values of each test when used for purposes of diagnosis. One may prefer other tests than those here included; but until wider use is made of diagnostic methods, this section deserves careful attention. Still finer instruments of diagnosis are described under "perception tests" and "tests for motor processes." Most of the test material in these sections can be used without special apparatus. The tests here given include tests for the rate of vocalization, for interpretation, for extraneous movements, breathing during oral reading, eye-movements, etc.

The last division of the book consists of a critical account of the remedial work which has been carried on by various teachers and investigators. Because of its summarization of the leading monographs and magazine articles on reading, this section will be valued by many busy persons who have become exasperated in their attempts to hold in mind the main currents of the recent movements in reading.

By presenting, first, an analysis of reading; second, an array of diagnostic instruments; and third, a large number of tested plans, Doctor Gray has provided a scientific book of the first order. The book should find a place in the activities of all practitioners in reading. It would have great value if used as a basis for discussion in a series of teachers' meetings.

WILLIS L. UHL

University of Wisconsin

BROOKS, SAMUEL S. *Improving schools by standardized tests.* Boston: Houghton Mifflin Company, 1922. 278 pp.

This book is an interesting and detailed story of how Superintendent Brooks, a district superintendent of schools in New Hampshire, utilized standardized mental and educational tests in the supervision of twenty-six rural, ungraded schools. The book begins with a brief description of the conditions existing within these schools when he became district superintendent. It gives a vivid picture of the inexperienced and untrained teachers, the poor system of grading, the scantiness and poor selection of materials of instruction, and the decrepit conditions of the buildings. It then tells in detail how Superintendent Brooks interested the teachers in the giving of standardized tests, how he trained them for administering the tests, and what use they made of the results. It shows how the results were used for classifying the pupils, for diagnosing the weaknesses in the instruction of individual children, for modifying instruction, and for measuring the progress of pupils, the efficiency of teachers, and the efficiency of special teaching methods. One chapter describes the application of intelligence tests to individual pupils, and another

chapter gives the reaction of pupils and teachers to the use of the tests. The last five chapters of the book constitute a description of the attempt of the superintendent and teachers to overcome the poor reading condition revealed by the testing. These chapters are rich in detail and set forth in clear manner a constructive program for the teaching of silent reading.

As pointed out by Superintendent Brooks in the preface: "The book is unique in at least two respects: (1) it is a narration of an actual experience rather than a theory, and (2) it describes in detail how a comprehensive, periodical testing program was planned and carried out, how the interested cooperation of the teachers was secured, and how the results of the tests were used to improve the schools." It is a book on the successful administration and utilization of tests. At the same time it is an excellent treatise on the methods of teaching reading. To the mind of the reviewer it represents an attempt to unite measurements and method. As Buckingham so well says, the book describes a "testing program carried to its logical conclusion." While the reviewer does not approve of placing so much dependence upon the single application of a standard test as Superintendent Brooks apparently did, he nevertheless unhesitatingly recommends this book to superintendents, principals, and teachers.

CLIFFORD WOODY

University of Michigan

WILEY, J. A. *Practice exercises in supervised study and assimilative reading.* Cedar Falls, Iowa, Iowa State Teachers College: J. A. Wiley, 1922. 112 pp.

Professor Wiley, in this little book, has added something decidedly worth while to our limited literature on Supervised Study and Teaching Children How to Study. Its contents are unique, concise, and practicable. It will prove a God-send to teachers who have awakened to the fact that in order to study efficiently children must be taught how to study and who feel the need of a "guide for directing the formation of efficient study habits."

In his preface the author says in part, "This book is constructed on the theory that skillful study involves the use of a great variety of specific study habits, each of which must be built up in conformity with the laws of habit formation. We do not acquire a habit by being told about the procedures involved, but rather by practicing those procedures over and over until they are perfected and reduced to the habit level. Likewise, in learning to study, pupils must first be shown how to use each of the better specific study procedures, and then they must be directed through sufficient practice exercises to habituate these study procedures."

It is significant that the author recognizes that efficient study (from books) involves many "specific study habits" each of which must be specifically developed if the pupil is to become an all around good student.

The book consists for the most part of directions to the teacher for conducting class exercises designed to develop these specific study habits. It is divided in three parts. Part I deals with the "General Problem of Rapid Thought Getting" from the printed page, Part II with "Specific Problems

in Comprehension'' in silent reading, and Part III with ''Organizing and Impressing Thought.'' Each part contains several chapters. Each chapter deals with some particular phase of good study ability, and contains definite directions for conducting several types of class drill designed to habituate the study procedure with which the chapter deals. In addition, the book contains directions for checking up progress in the formation of efficient study habits by means of standardized tests. There is also an introductory chapter which fits the book for use as a text for a correspondence course in supervised study.

On the whole, the book is one of the best aids we know to the solution of the difficult problem of teaching children how to study.

S. S. BROOKS

District Superintendent Winchester, New Hampshire

MEYER, MAX F., *A brief manual of psychology demonstrations.* Columbia, Mo.: Missouri Book Company, 1922. 108 pp.

This book is designed to furnish illustrative materials for ''The Psychology of the Other One.'' It ''will help that student who uses the author's introductory text, to apply his own hands to the demonstration of certain not quite ordinary experiences with other people (not with his consciousness).''

The exercises include many of the typical and usual experiments of the psychological laboratory—exercises in reaction-time, estimation of weights, space-perception, etc. However, the treatment of results and the explanatory material are not at all conventional and undoubtedly justify the author's contention that ''these demonstrations illustrate and thereby further clarify the theories of which the text is composed.''

E. H. CAMERON

University of Illinois

MEYER, MAX F., *The psychology of the other one.* Columbia, Mo.: The Missouri Book Company, 1921. 439 pp.

In ''The Psychology of the Other One'' we are introduced to the subject in a highly dramatic way. Robinson Crusoe has just discovered Friday and is concerned with the problem of how to use him. The Devil appears in the guise of a missionary and tries to persuade Crusoe to begin his study with the assumption that Friday has a soul. Crusoe, being thoroughly up-to-date in matters of psychology, has no difficulty in refuting the evil one even when he changes his appearance to that of a philosopher. He is firm in his resolution to restrict his interest in Friday to what his own sense organs, or those of others, will reveal; for only in this way can he measure Friday, and there can be no science without measuring.

As one reads the succeeding chapters, one can scarcely avoid the conclusion that the Devil has lived up to his reputation. He has got Crusoe so excited about denying the value of the soul that he passes by some very obvious and important aspects of Friday's nature. Furthermore, he proceeds at once to forget his formula of dependence upon sense organs for his observations.

Chapter II introduces us to the study of an imaginary lump of living tissue something like a snail. It is called a snail for convenience but is much simpler than a real snail. This imaginary animal's activities are explained by an imaginery nervous system so simple that its functions can easily be described in terms of physico-chemical changes. In this way we learn of reflex action.

In Chapter III we are introduced to the conception of "concerted action" by the study of the behavior of a real jelly-fish with a real nervous system. But to understand the "concerted action" of the jelly-fish it is necessary first to "imagine the analogous case of eight leaky places on a water pipe. Suppose the frequency of the dropping at each place to be about the same, say, one drop a second. But even then we could not expect all the eight drops to fall simultaneously; as we say in physics, we could not expect that there be no phase difference among the eight drops. The eight drops would probably fall from the eight leaks in a quite irregular succession." Apply this analogy to the nervous system of the jelly-fish and you have an explanation of its "concerted action." Friday, too, exemplifies concerted action. For example, if an insect lights on Friday's nose this is what happens: first, he twitches his nose; second, if the insect has a good hold on his nose he shakes his head; third, if this does not remove the insect, he brushes it off with his hand. Leaky water pipes explain Friday's behavior in this instance just as they do the concerted action of the jelly-fish.

After such an introduction it becomes possible for the student to imagine that he has "received from the Creator an appointment of serving as his assistant, to have received the task of constructing for the benefit of a given animal a nervous system which will fulfill its needs." The remainder of the book is largely devoted to this task of architectural design for a nervous system which will explain Friday's behavior. The task is performed in a highly elaborate and ingenious fashion. The Other-one's attention, space-perception, talking and writing, walking in his sleep, etc., are all described and explained by this method. The book ends with one last shot at the Devil, who does not, however, appear this time in visible form. "Our interest in other human beings as souls positively is highly dangerous for human society. . . . The most cruel acts of man against man are those committed in the name of man's mysterious soul. . . . A rational, a humane treatment of its individual members is to be expected much more by a society which regards itself as a group of organisms than by one which regards itself as a mysterious collective soul."

E. H. CAMERON

University of Illinois

HOLLEY, CHARLES ELMER. *The teacher's technique.* New York: The Century Company, 1922. 378 pp.

According to its author, this book has been prepared "for the large group of teachers who are compelled to attack the teaching problems of the grammar grades and high school with very little training to help them." The foregoing probably constitutes the reason for the wide range of topics treated in the book and also for the familiar, descriptive style employed.

Technique is not defined, but it is used to mean *knowing how and when to perform all the duties of a teacher.* The author's intention is that people with little training for teaching will read, and possibly study, the various chapters of the book in connection with their work in teaching. To assist such students, the author has provided summaries and exercises for each chapter and also a detailed outline of the whole book. These devices make the author's positions clear and exact—so clear and so exact that, were it not for the author's clearly expressed purpose, he might be charged with being dogmatic.

Just how a lone untrained teacher would get on with the book is uncertain. It is, however, certain that a group of teachers in a study center, under the direction of a superintendent or extension teacher, would find the book profitable. Such group discussions, if current teaching problems were made prominent, would yield understanding, appreciative insight, and increased skill.

The wide range of topics and the limitations of the size of the book are presumably responsible for such startling statements as these:

"The page assignment is the ideal one toward which a teacher should strive." (p. 76)

"The recitation is intended primarily to serve as a means of ascertaining whether the pupils have performed the tasks set for them in the assignment." (p. 78)

"The task of imparting information is the most important one that confronts the teacher." (p. 112)

"The real psychological basis of emotional teaching (?) is the radiation of feeling attitudes from teacher to pupils." (p. 163)

"This radiation of emotional attitudes is the keynote of the method needed for imparting moral standards." (p. 220)

"The term, review, means in reality a new view." (p. 182)

The book has merit. In a sense, it is a pioneer in the field of what constitues teaching technique. To have this field tentatively defined is a distinct service to the teaching profession. Then, too, to present within a chapter the essential things that have been written in several volumes on such subjects as teaching children to study, developing individuality, outcomes of teaching, is a distinct service to all of us, including the beginners without preparation. The book is an attempt to state the whole round of the teacher's work today in much the same way that Page's "Theory and Practice of Teaching" stated it more than a half century ago.

JOHN A. H. KEITH

State Normal School, Indiana, Pennsylvania

BURDGE, HOWARD G. *Our boys. A study of the 245,000 sixteen-, seventeen-, and eighteen-year-old employed boys of the State of New York.* Albany: J. B. Lyon Company, 1921. 345 pp.

There have been many statistical studies in the educational field during the last few years, not a few of which have proved invaluable in helping to establish more scientific methods of administration. But if we limit considera-

tion to the specific type of statistical studies represented by this publication, probably the study entitled "Our Boys" stands out supreme.

Its claim to superiority is based upon three things:

1. The unusually large number of records obtained. (The investigation had to do with no fewer than 245,000 sixteen-, seventeen-, and eighteen-year-old employed boys in New York state.)

2. The breadth of the survey. (Forty-two distinct answers of significance for social workers and students of economy as well as educational administrators were obtained from each boy.)

3. The highly reliable nature of the figures compiled. (This reliability was made possible by the Governor's edict which required that all information be recorded by public school teachers in public school buildings.)

The reviewer cannot refrain from expressing the hope that the same or equally good machinery might be frequently employed in connection with investigations having to do with social and economic welfare.

Since the author of "Our Boys" has been satisfied with collecting in intelligible form this vast body of information, leaving the interpretation of the same largely to others, and since so many important facts are revealed therein, it remains merely to point out such of these as promise to have greatest significance for readers of this magazine.

1. Six-sevenths of all sixteen-, seventeen-, and eighteen-year-old boys in New York state are out of school.

2. Only four boys out of five claim the father as guardian.

3. Twice as many fathers as mothers were reported dead.

4. Five percent of the boys have neither father nor mother as guardian.

5. Over 65 percent of these boys remained in school one or more years beyond the compulsory age limit.

6. Twenty-five percent left school as soon as legal requirements were met.

7. The oldest boys in families make more rapid progress in school than their younger brothers.

8. In Greater New York the proportion of boys who complete the eighth grade is approximately 50 percent greater than that of boys employed on farms.

9. The vast majority of these boys left school because they "wanted to go to work" and not because they were obliged to.

10. Mathematics is the best liked study.

11. English is the least liked study.

12. Less than 10 percent of the boys attend night school; and over 60 percent state that they do not wish to attend such schools.

13. Less than 2 percent of the boys are assisted by schools, churches, and employment agencies in getting employment.

14. Over 40 percent spent less than four and one-half months on their last job.

15. There is a distinct, though not marked, correlation between fathers' and boys' occupations.

16. There is a high correlation between grades completed and occupation entered and followed. Those who leave school on or before completion of the

eighth grade enter, for the most part, the industrial trades and occupations. Those who complete one or more years of high-school work enter the clerical and retail business occupations to a greater degree than the members of the other group.

17. There is but slight correlation between present and desired occupations and best and least liked studies.

By no means the least valuable part of this publication is the tabular material from which the numerous correlations were derived. Of particular interest to the "doubting Thomas" type of superintendent is the complete vindication herein published of the random sampling method of interpreting data. It seems that a group of skeptics in Buffalo somewhat clandestinely tested the validity of this method by laboriously checking every single item of information regarding the 6,468 local boys and comparing the results obtained with those obtained by means of the random sampling method. The published comparisons provide strong and unequivocal support for the latter method.

While it would be unsafe for any local administrator to seek to apply in full the above conclusions to the solution of his problems, it is safe to say that many of these conclusions doubtless have almost universal application. The chief value of the study, however, (outside of the direct contributions cited) rests in the advertisement of the fact that we have in our public school system the machinery for carrying on investigations of real significance to society and that such investigations can be managed more effectively through this medium than through any other medium heretofore used.

D. J. MacDonald

Ohio State University

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

Doctor Dorothy Hazeltine Yates, formerly a member of the Stanford University Research Staff for the study of gifted children, now teaches psychology at the State Teachers College, San Jose, California. She is also assistant director of research.

The Edward E. Babb Company, Boston, announces the Beatley I.Q. Slide Rule, and states that it has been used for more than a year with full satisfaction at Harvard University School of Education. The instrument is apparently accurately designed. It is made of hard wood in a form convenient for use. Anyone who has a large number of I.Q.'s to figure will find this a labor-saving device.

We have recently received a prognostic test of stenographic ability arranged by Doctor Elmer R. Hoke of Lebanon Valley College, Annville, Pennsylvania. Mr. Hoke has also devised a measuring scale for the knowledge of Gregg shorthand on the general plan of Ayres or Ashbaugh spelling scales. This scale is published by the Gregg Publishing Company.

The school children of America saved more than $5,775,000 in the school year 1921-22, according to figures compiled by the Savings Bank Division of the American Bankers' Association. The result of their thrift is an increase of 40 percent over the preceding year, and 100 percent more than the total of the year 1919-20.

One city, Louisville, Kentucky, reports that 98 percent of its 27,717 children have savings accounts. Wilkes Barre, Pennsylvania, claims first honors on the basis of individual deposits averaging $7.54.

We are glad to note that the Board of Education of Los Angeles is issuing Volume II of its *High School Research Bulletin* in printed form. The bulletins last year were outstanding in character. Number 1 of Volume II gives promise that this year's issues will be equally important.

The report of Doctor Bobbitt's work on the curriculum in Los Angeles last year has been issued in Number 20 of the Supplementary Educational Mono-

graphs published by the University of Chicago. Doctor Bobbitt will spend January, February, and March, 1923, in Los Angeles continuing the curriculum study.

The Board of Education of Pittsburg has issued in its *Bulletin Number 4* of the Department of Research and Measurement a course of study in geography. The work was done by a committee under the direction of Mr. Guy, director of the Bureau. Type studies and extensive references are given for the work in grades III to VIII, inclusive. The bulletin also contains a parallel course based upon the textbook for grades IV-A to VI-A, inclusive. Individual schools are left free to choose between the two courses. The bulletin is a veritable mine of valuable suggestions for teachers of geography.

Persons interested in part-time education should secure monograph Number 2 of the 1921 series, published by The Vocational Education Association of the Middle West. Orders should be addressed to L. W. Whalstrom, 1711 Estes Ave., Chicago. This pamphlet contains an article on each of the following topics: legal standards, standard methods of securing suitable subject matter, and problems in continuation schools for girls.

The *University of Kansas Bulletin*, Volume XXIII, Number 1, contains a survey of the schools of Lawrence, Kansas. It was made in the school year of 1920-21 under the direction of Mr. F. P. O'Brien, director of the Bureau of School Service. The printing of the bulletin was delayed in order that at least one of the results of the survey could be reported, i. e., the successful campaigning for the issuance of bonds for a new building up to the full legal limit of the city's capacity. The report is in three sections: (1) the curriculum needs for secondary education, (2) the school-building needs, and (3) financing the school-building program.

The school and the church in Dublin (Ohio) have joined forces in the publication of a monthly journal for "the promotion of the religious, moral, and social welfare in community life." The journal contains a large amount of information about the schools which is in this way brought into the homes of the people of that little community.

On November 9 and 10, 1922, Kansas University held a conference on Educational Measurements. The University hopes to make this conference an annual affair. The principal speaker was Doctor B. R. Buckingham who addressed the conference on the evening of November 9 on "The Ability of Pupils to Solve Verbal Problems in Arithmetic." He also addressed the conference on November 10 on "A Common-Sense Testing Program," and on "Follow-up Work with Special Reference to Reading." We hope that Dean Kent and his colleagues will find enthusiastic support among the school people of Kansas for this University conference.

At Youngstown, Ohio, an experimental attack upon the problem of improving the ability of children in arithmetic has been inaugurated. Doctor

Henrietta V. Race, director of research for the board of education, reports that the approach to the problem is being made through a consideration of the psychology of arithmetic. Considerable work of this sort has been done in reading as is attested by the numerous reports now coming from the principals. As a practical school man remarked a few weeks ago, "We need the same thing in arithmetic." Perhaps Youngstown will supply the need.

Assistant Superintendent Albert S. Cook of Maryland has just issued a helpful bulletin on silent reading. He takes the position that testing should precede teaching—the same position which is being taken with reference to other school subjects, notably with reference to spelling. The bulletin gives information as to the choice of reading tests, and takes up the method of training for the improvement of both speed and comprehension. Norms for silent reading as they have been developed in typical Maryland counties are printed at the close of the bulletin. If this pamphlet is widely circulated in Maryland it can scarcely fail to be helpful to teachers of reading.

Superintendent M. G. Clark of Sioux City, Iowa, has issued an interesting pamphlet on "Making High School a Success." It embodies "a discussion of the atmosphere, the out-of-school environment, and the attitude of mind which must be maintained to guarantee to a pupil a measurable degree of success in public high-school work." In answer to the question, "Does education pay?" the bulletin offers the emphatic answer, "It *does*." It discusses the fundamentals of high-school success and compares the advantages and disadvantages of public and private high schools. Much of the pamphlet is addressed to the parents of public school pupils.

With further reference to the advantages of high-school training the Harter School Supply Company prints a pamphlet in which every school official should be interested. Its appeal is directed to parents and to actual or prospective high-school pupils. It presents in a forcible and readable way the social and financial opportunities which secondary education confers. Not the least interesting feature of the bulletin is the set of cartoons which are interspersed with brief statistical statements and editorial comment.

Doctor Luella C. Pressey contributes the following story, illustrative of the fact that most schools fail to adjust materials and methods to the needs and abilities of pupils: Tom brought home a bad report card; whereupon, Tom's father went "up in the air." "Why," inquired the disgusted parent, "why aren't you at the top of the class?" Tom fumbled around in his mental *repertoire*—which, after all, didn't appear to be meager—and finally dug up this reply: "What's the use, father, of me being at the top instead of at the bottom? They teach the same things at both ends."

The Bureau of School Service and Research of the University of Kansas is conducting a state cooperative study of high-school instruction. Two hundred cities in that State signified their readiness to participate in the

study. The Bureau has undertaken to score and tabulate results for the schools which felt unable to participate otherwise. The program of testing to date includes the Terman Group Test, Thorndike-McCall Reading, Brigg's English Form, and Hotz Algebra. It is intended that this study will provide a continuous program of testing to extend over perhaps a two-year period until many of the high-school subjects have been included and some tests repeated.

F. P. O'Brien, the director of the Research Bureau of the University of Kansas, has recently completed a survey of instruction work in the junior and senior high schools of Lawrence, Kansas. The survey employed educational and mental tests for judging the results of instruction. Nine of the former and two of the latter were employed in both the junior and senior high schools. The other items treated in the survey were courses of study, supervision of instruction, teaching staff, and costs of instruction. The report will be printed for distribution later.

The Lawrence School Survey has just appeared in the form of a 100-page printed bulletin of the Kansas University Extension Division. It deals with the curriculum needs of the high school, a 30-year school-building program and the financing of such a program. The survey was conducted by the Bureau of School Service and Research, under the direction of F. P. O'Brien. The city has already constructed a modern senior high-school building as the first step in carrying out the building program recommended in the survey report. Several other cities in the state have since authorized similar surveys by the same bureau.

A study recently completed as a master's thesis at the University of Kansas shows that for four high schools in Kansas City, Mo., the students who are leaders or active participants in the organized activities of these high schools rank above the non-participants both in school work and in their Army Alpha test scores. Eligibility regulations, so far as any existed, would have restricted a negligible percent of the non-participants from being active in these high-school organizations.

Monograph Number 4 issued by the Board of Education of Berkeley, California, is entitled "The Use of Mental Tests in School Administration" and is written by Doctor Virgil E. Dickson, director of the Bureau of Research and Guidance. The monograph contains a detailed report of the mental testing done in the Berkeley schools with reports concerning the reliability of tests and their use in prognosis. The following excerpts from the summary and conclusions are worth passing along:

1. On the basis of a follow-up for two years, it is shown that approximately seventy-five percent of the kindergarten children with an I.Q. of below 90 failed to make normal progress, and that none made rapid progress; while of those having an I.Q. of 110 or above, about one-third made rapid progress

and less than ten percent made slow progress. The figures for the first grade most emphatically justify the conclusion that mental tests in the kindergarten have a definite prognostic value.

2. A study of the school records for four and a half years of a group of children who were given intelligence tests in the first grade showed that the prediction as to school progress—rapid, normal, or slow—has proved true in seventy-eight percent of the cases. In eighteen percent of the cases the failure of the prediction was explained on the basis of absences, illness or environmental hindrances.

A study of the relation of some physical defects to achievement in the elementary school by Doctor J. N. Mallory has been issued by the George Peabody College for Teachers as Contributions to Education, Number 9. The study was based on physical examinations in charge of a registered nurse, and the scores made on the following tests: Cleveland Survey Arithmetic, Buckingham Extension of the Ayres Spelling, Hahn-Lackey Geography, Freeman Writing, Monroe Silent Reading, the Nassau County Supplement to the Hillegas Composition, the Illinois General Intelligence, and Holley Picture Completion.

The conclusions are in part as follows: physical defects are directly associated with low achievement scores; physical defects constitute one cause of retardation; and retardation constitutes one cause of backwardness in achievement. Consequently, physical defects act both directly and indirectly to cause backwardness in achievement. All defects do not constitute handicaps to progress in the same degree. In this respect they should be properly ranked in the following order: nasal obstructions and mouth breathing, defective teeth and resulting maladies, defective hearing, defective tonsils, and defective eyes which includes defective vision as well. The last named, however, seems to be only slightly associated with low scores and in some cases the association is negative.

The *Survey Bulletin* of the schools of Bucyrus (Ohio) for November, 1922, deals with the use of achievement and intelligence examinations in that school system. Superintendent Paterson makes three surveys of his schools each year; and attempts to obtain something really significant from each survey. Data are interpreted, conclusions extracted, and remedial efforts gotten under way. The bulletin presents a large amount of data on the various tests given to all grades in the schools, and the final summaries of achievement are shown in relation to rank in brightness or the sizes of scores on the intelligence examinations. Superintendent Paterson evidently attempts to impress his teachers with the fact that, in general, achievement ranking should be identical with brightness ranking and that any departure there from is indicative of the strength or weakness in the teaching situation.

The following data from Assistant Superintendent Joseph A. Baer of Cuyahoga County (Ohio) may be of interest to those who are making successive measurements with intelligence tests:

COMPARISON OF I.Q.'S ON SUCCESSIVE TESTS ILLINOIS EXAMINATION

Grade	Number of cases	First Test Dec., 1921	Second Test Oct., 1922
III	16	100.9	104.3
IV	12	104.8	105.3
V	11	113.8	112.2
VI	14	115.2	118.4
VII	10	116.9	120.0
Total	63		
Average		112.7	113.3

Of course, these data are not sufficient to be conclusive but there is a sufficiently high correlation ($r = 0.86$ based on individual scores) to suggest that such measurement is likely to be a rather stable indication of the situation on any measurement. Superintendent Baer says: "It would seem that the Illinois Test measures some quality which is relatively constant with respect to age. I have not heard anyone assuming that this quality is the whole of that elusive something called general intelligence. The test does measure well as far as it goes, and it goes far enough to predict pretty definitely the ability of a pupil doing regular school work. After all, that is the important consideration just now."

Some of our readers may remember the first report of the Committee on Research of the Upper Peninsula Educational Association (Michigan), published in 1913. A new report by the same committee has recently come from the press. The present report covers the situation in the Upper Peninsula on the following points: rural school conditions, cardinal principles of secondary education, and the financial status of city-school systems. The report on rural-school conditions shows that 80.5 percent of the teachers are graduates of four-year high-school courses. Only 0.02 percent of the teachers are without any professional training; although 52 percent have had but six weeks which is the minimum amount required by law. Sixteen percent have had more than 18 weeks of professional training. In the spring of 1922, 26.4 percent of the rural teachers were completing their first year; 15.8, their second year; and 19.8, their third year. The median teaching experience of rural teachers is approximately two and one half years, and the median age is twenty. The rural teacher has, on the average, six grades with a median of twenty-three classes daily, though some reported as high as forty-seven.

Sixteen of the twenty-six high schools require physical training; eleven have a form of medical inspection; and only one has a dental clinic.

Superintendent William L. Ettinger in a recent address delivered before the associate and district superintendents, examiners, directors, and inspectors of the New York City schools showed that on the average each one thousand children who had entered school in that city eight years before were distributed as follows: graduates, 139; VIII-B, 260; VIII-A, 288; VII-B, 185; VII-A, 87; VI-B,

30; VI-A, 9; V-B, 2. The percent of promotions was distributed promiscuously throughout the grades and seemed quite independent of chronological age. For example, among 47,000 pupils in the low-first grade the ages ranged from less than 5 to 18, and of the 29,000 pupils in VIII-B the ages ranged from 10 to 18. On the other hand, the distribution of twelve-year-old children was found to be from low-first to high-eighth. Forty-three percent of the pupils were found to be of normal age; 26 percent, under age; and 31 percent, over age. These data are on the basis of a span of a year and a half for each grade. Examining the situation from another angle, it was found that among the 30,000 children in the high-eighth grade (admission to which normally assumes 15 terms' work) three had been in school but 6 terms; six, 7 terms; eight, 8 terms; while, at the other extreme, four had been in school 24 terms; twenty-four, 23 terms; fifty-two, 22 terms. Among the 18,000 who were in their sixteenth term in school, it was found that one was in high-second grade, two in the low-third, eleven in the high-third, and so on until twelve were found in the low-tenth, and five in the high-tenth.

Variability in achievement was found to have a similarly wide distribution. For example, in spelling, among high-eighth-grade pupils, 30 percent fell below the low-eighth-grade average, 13 percent below the high-seventh average, and so on down to 4 percent who were below the low-fifth.

Mr. A. O. Heck, director of the Bureau of Research at Akron (Ohio), has reported the following activities of his Bureau for the years 1920-22:

For the year 1920-21.

I. Educational Tests. A survey was made of the entire system in grades IV-B to VIII-A, inclusive, in both October and May in reading and arithmetic, and in October in spelling and writing.

II. Practice Tests in Arithmetic. The Courtis Standard Practice Tests were used in three schools, and the Studebaker Practice Exercises in two schools. The results were checked against two control schools.

III. Mental Group Tests. The National Intelligence Tests were used throughout three schools from IV-B to VIII-A, inclusive, and the Pressey Primer Scale in the same number of schools in lower grades.

IV. Teachers and principals were trained in the use of tests. This was accomplished through meetings held with the Director, and the normal students were given courses in "Measurements."

V. Age-Grade study.

For the year 1921-22.

I. Educational Tests. A survey was made of the entire system in grades IV-B to VIII-A, inclusive, with Courtis Supervisory Tests in Arithmetic in September, January, February, and May. An experiment was carried on in reading for the purpose of improving both rate and comprehension. Experimental groups were taught by teachers who took a course in measurements in which reading was emphasized, and the method of teaching set up. Control groups were taught by teachers of the same building who were judged by their principals to be of nearly the same ability as those teaching the experimental group. Progress was checked by the use of the Burgess Reading Scale.

II. Practice Tests in Arithmetic. Courtis Practice Tests were used in five schools and Studebaker, in three.

III. Mental Group Tests. One school sectioned its pupils in each grade from I-B to VIII-A, inclusive, on the basis of mental-test results. Two

schools were tested at the end of the year and sectioned for the following year. All pupils in the eighth grade were tested each semester so that reports of tests were available to the high-school principals. The purpose was to secure data which would possibly enable teachers more wisely to guide the pupils entering high school in their choice of studies.

IV. Teachers and principals were trained through meetings held from time to time, and through classes in extension courses in educational tests and measurements.

V. Age-Grade study.

The above outline of work of the Bureau at Akron is given for the benefit of those who may not be acquainted with the lines of activity engaged in by such Bureaus in city school systems. The work at Akron is not as extensive as in some places, but the report of it is decidedly impressive.

A Note on Arithmetic

We all agree with the wisdom of Thorndike's insistence on drill upon the specific connections which are weak in contrast to indiscriminate practice. In actual classroom teaching or supervision it is often baffling to determine which connections are weak. The following is an example of determining the weak connections.

Consider the following four examples all done incorrectly by members of a fifth-grade class in the University Elementary School, Iowa City, Iowa.

a	*b*	*c*	*d*
68798	68798	65164	64356
79856	79856	51948	78493
68597	68597	19485	46532
217241	197251	136587	189371

The correct answers are:

217*2*51	*2*17251	136*5*97	189*3*81

It is obvious that these children form the right connections between the numbers themselves, i. e., they know that $7 + 6 = 13$, that unseen $13 +$ seen $8 = 21$, etc., (from units column, example *a*). Every mistake is one of carrying. Three times they failed to carry at all. Once they failed to carry 2 but did carry 1. In 16 chances to carry, they did so correctly 12 times, once the carrying was wrong, 3 times they forgot. The "carrying" habit is weak and that is where special drill is needed. This analysis took just three minutes, not an undue amount of time for a classroom teacher to spot the place for drill, especially since directed drill saves everyone's time in the end.

F. B. KNIGHT

University of Iowa

Testing Program for Schools of Michigan

Sixty-seven school systems in various sections of Michigan united during the second week of October in a cooperative testing program, directed by the Bureau of Educational Reference and Research of the University of Michigan. The extent of the program is clearly shown by the following tables.

TABLE I. NUMBER OF TESTS USED IN TESTING PROGRAM

	Tests	Number
Part 1.	Elementary Grades:	
	Monroe Silent Reading Tests, Form 1	51,000
	Courtis Supervisory Tests in Arithmetic, Forms 3-4	72,000
	Woody McCall Mixed Fundamentals, Form 1	30,000
	Spelling Test for all grades, prepared by the Bureau based upon the Ayres-Buckingham Spelling Scale.	
Part 2.	High-school English (A study of the improvement resulting from the first year of instruction in high-school English):	
	Van Wagenen Reading Scale in English Literature, Scale A	6,500
	Kirby Grammar Scale, Form 1	8,000
	Briggs English Form Test, Alpha	7,500
Part 3.	High-school Latin—(An evaluation of the special phases of teaching Latin in cooperation with the American Classical League):	
	Carr's English Vocabulary Test, Form A	4,500
	Uhlman-Kirby Latin Comprehension Test, Form 1	1,600
	Godsey's Latin Diagnostic Composition Test, Form 1	1,600

TABLE II. SCHOOL SYSTEMS PARTICIPATING IN PROGRAM

	Parts[1]		Parts[1]		Parts[1]
Adrian	1, 2, 3	East Lansing	1, 2, 3	Manistee	1, 2
Albion	1, 2, 3	Farmington	1, 2, 3	Marine City	1, 2
Allegan	1	Ferndale	1, 2	Marysville	1, 2, 3
Alma	1	Flint	1, 2, 3	Mich. State Normal	1, 2, 3
Ann Arbor	1, 2, 3	Fenton	1, 2, 3	Midland	1, 3
Algonac	1, 2, 3	Goodrich	1, 2, 3	Morley	1, 3
Bad Axe	1, 2, 3	Grand Rapids	1	Mt. Clemens	1, 2, 3
Baraga	1, 2	Grayling	1, 2, 3	Mt. Pleasant	1, 2, 3
Battle Creek	1	Greenville	1, 2, 3	New Troy	1, 2
Bellevue	1, 2, 3	Gwinn	1, 2, 3	Otsego	1, 2, 3
Berlamont	1	Harbor Beach	1, 2, 3	Paw Paw	3
Berrien Springs	1, 2	Hastings	1, 2, 3	Pontiac	1, 2, 3
Big Rapids	1, 2, 3	Highland Park	1, 2, 3	Port Hope	1, 2
Birmingham	1, 2, 3	Holley	1, 2	Saginaw, W	1
Boyne City	1, 2, 3	Hopkins	1, 2, 3	Sault Ste. Marie	1, 2, 3
Britton	1	Howell	1	Stanton	1, 2
Cadillac	1, 2, 3	Hudson	1, 2, 3	Tawas City	1
Carson City	1	Ionia	1	Vicksburg	1, 2, 3
Coldwater	1, 2, 3	Ironwood	1, 2, 3	Wyandotte	1, 2, 3
Dowagiac	1	Iron River	1, 2, 3	Yale	1, 3
Durand	3	Jonesville	1, 2, 3	Ypsilanti	1, 2, 3
Dundee	1, 2, 3	Kalamazoo	1, 2, 3	Zeeland	1, 2, 3
		Lakeview	1		

[1] The numerals 1, 2, 3, after the names, represent the parts of the program—elementary, English, and Latin, respectively, in which the schools participated.

According to the provisions of the program each school system will forward its tabulated results to the Bureau where general summaries will be compiled and published in its regular bulletin.

Some Fallacies in Testing Intelligence Tests

In the March, 1922 issue of this *Journal*, Henmon and Streitz presented an article entitled "A Comparative Study of Four Group Scales for the Primary Grades." These scales are: Pressey Primer, Haggerty Delta 1, Dearborn Series 1, and Myers Mental Measure. The authors compared and appraised these four tests chiefly on the basis of four criteria. Assuming for argument that these are valid criteria, we shall examine the procedure and the explicit and implicit inferences by these authors.

Correlation with teachers' estimates.—"Aside from the Pressey test in the second grade of school A, and the Myers test in the first grade of school B, the correlations are high and do not differ greatly," they say. But they do not say how many children were in the first- and second-grade classes tested in each school, and there were only 50 children altogether in the first-grade group and in the second-grade group, respectively. In spite of this fact the authors draw rather definite conclusions, which must suggest to the average reader of their study that these data are representative of all children. Of course probable errors are given, but a probable error can have little significance unless there is a guarantee that the cases were a wholly random selection. It were almost impossible to pick out two entire school classes of a given grade anywhere that would be representative, say, of 10,000 children randomly selected from children of that grade. If these authors had picked out four other classes, who knows but that the relative correlations for the four tests would have been entirely different.

We do not know whether any of the children tested were of foreign parentage or not, but since none scored very low on the Haggerty test which presupposes familiarity with some school symbols, one may infer that all of these children were fairly familiar with English. But in the average American class of school children, especially of the first and second grade, there are children who can scarcely understand or speak English. Among such children these four tests would give, according to this criterion as well as by the other three criteria, very different ranking of the tests, perhaps, from that found by Henmon and Streitz.

Discriminative capacity.—"Other things being equal, that test is most useful which segregates most clearly and unambiguously different age or grade groups," they say; and then, ignoring the age factor, they study overlapping purely on grade basis. How does the reader know whether or not there are more over-aged children in the second grade than in the first, that the larger overlapping by the Myers and Pressey tests is not what the probable age discrepancy would warrant? Granting no age discrepancy, and that for ages these tests ranked for overlapping as they do for grades, even then by what authority of fact do Henmon and Streitz say that the less overlapping the better. How do we know whether the overlapping of actual intelligence, if correctly measured, would not even be greater than that found for the Myers Mental Measure, the test which showed for them the greatest overlapping? It may be that a six-year-old child does not differ in native capacity from the seven-year-old child nearly so much as any verbal test has suggested. Maybe

the apparent difference has been due, in a large measure, to what was indicated by those elements in intelligence tests (including the Stanford-Binet) which are most affected by school experience.

On page 193 the authors proceed to compare medians and the differences between the medians of the first and second grades for each test. But this procedure is wholly unsound since, in the first place, the ranges of the four scales and the successive units of measure on each scale are not at all commensurable. Worse still, the Haggerty, Dearborn, and Pressey tests are designed for the first three grades only, whereas the Myers Mental Measure is designed for all ages and school grades. Because of its wide range of application the Myers test could not be expected to show so great difference between the medians of successive grades as would those tests designed for very limited range of ages and grades. Assuming, then, for argument that the Myers Mental Measure is least discriminating between the first two grades, of the four tests compared, it should still be noted that the authors start their thesis on this subject with: "Other things being equal." The impression left is that other things are equal. If they wished to leave a correct impression, one might question why they did not point out this unique feature of the Myers Mental Measure which, because of its universal range of application, admits of the study of overlapping not only of the first two or three grades but of these three grades with higher grades. The break of the Dearborn, Pressey, and Haggerty tests at the end of the third grade forbids one's finding by them the overlapping of the third and fourth grades, of the first and fourth grades, of the second and fourth, of the third and fifth, and so on.

Character of the distribution curve.—Can any one of scientific habits consider seriously the relative distributions of these four tests by 50 cases? Assuming again for argument that the number of cases was large enough for consideration, even then the conclusions arrived at by the authors are not borne out by the curves and tables upon which these conclusions are based. In spite of the facts to the contrary, the authors says (p. 194): "but an inspection of Figures 1 to 4 and Table I show fairly close conformity to the normal curve except in the case of the Myers test." Just why "except"? Let the reader paste a screen over the labels of these graphs and present them to any group of students familiar with the "normal curve"; or let him present a copy of a "normal curve" to a group of intelligent laymen and ask them which pair of graphs conform most closely to this model, and he will find the Myers Mental Measure ranking more often first than last. Apparently the authors made up their verdict by a kind of visual average of the two graphs in each instance. Since they made no attempt to compare the distribution curves for each grade why did they not assist in this visual-averaging by merging the two graphs into one, that is, by merging first- and second-grade graphs for each test? If the reader will take the trouble to do that he will find that the Myers Mental Measure looks about the best instead of the worst. If then, he will turn to "Measuring Minds" by the writer and Mrs. Myers, pages 17-21, he will find for over 5,000 cases, approximately normal distributions for every age and school grade, including college students.

Administration and scoring.—Although the authors have failed on three counts to give conclusive proof of the relative validity of the four tests, they assume that the Myers Mental Measure was knocked out in the third round, for they wholly ignore it in the fourth. At this point the average reader of their article must have assumed that the Myers test deserves no longer to be considered by him as a possibility for use.

Henmon and Streitz seem to evaluate this last point on general impression for they write: "According to the examiner's reports, the Pressey test ranks high in these respects since it can be given in twenty-five minutes and scored in three minutes. The Haggerty test also ranks high as it can be given in thirty minutes and scored in five minutes. But they seem disturbed over the longer time required for the Dearborn test and think it "fatiguing and even boresome to many of the children." However, they do not tell us the order or time in which these four tests were given; neither do they suggest relative weights of the four criteria set up by them. If the umpires had wished to extend to the Myers Mental Measure the sportsman's consideration, they might have noted that this test can be given in less than half an hour and scored in about two minutes; and that it can be used as a continuous scale on an entire school population.

The shortcomings of the study of Henmon and Streitz are all the more serious because they implicitly suggest at the outset that they are going to furnish a rather conclusive guide for the selection of intelligence tests for the first two grades, for they say: "Departments of education and psychology are importuned for advice and counsel as to the validity, reliability, and relative merits of such scales as are in the market." The undiscriminating reader certainly can have come to no other conclusion than that these authors have furnished such "advice and counsel."

GARRY CLEVELAND MYERS

Cleveland School of Education

Educational Research Association

(FRANK N. FREEMAN, *Secretary and Editor*)

NOTE ON THE RELATION BETWEEN SPEED AND ACCURACY OR QUALITY OF WORK

Frank N. Freeman, University of Chicago

The problem of relationship between speed and quality is one on which there has been a rather wide divergence of opinion, and on which the evidence has not been particularly conclusive. Perhaps it is safe to say that the trend of opinion now is that there is positive correlation between speed and quality of work.

A rather well-known investigation on this point was that made on the subject of reading in the Cleveland Survey. By means of the ninefold grouping it was made clear that there was some correlation between rate of reading and comprehension. The correlation, however, was not high. Some years ago a graduate student in the Department of Education, University of Chicago, gave the Burt Reasoning Tests to a group of high-school students. The time allowance was sufficient to allow all of the students to finish, and the time taken by each student was recorded. The correlation between the score, which in this case was independent of the rapidity of the work, and the speed of the performance was calculated and was found to be zero. This seemed to indicate that for some types of performance at least, there is little or no relation between speed and quality of work.

In an effort to get additional information upon this problem a study was made of the relation between speed and score in two standardized tests on a class in mental tests during the summer quarter of 1922. There were 114 students, largely graduate students, from whom complete records were obtained. The class was given two tests upon the content of the course. One of these was a multiple-answer test, and the other a completion test. The rank order of the speed of performance in tests was determined by keeping a record of the order in which the papers were turned in. The first question to be asked is whether the order of completion is any indication of characteristic rate of work of the individuals. In order to determine this, the correlation between the order of finishing in the two tests was determined and found to be $0.50 \pm .05$. This indicates a fair degree of reliability of this measure. The reliability of the tests themselves is indicated roughly by the correlation between the two tests which was $0.55 \pm .047$. Since the content of the two tests was in part different this indicates again a fair degree of reliability. The relation between speed and quality of performance was measured by finding the correlation between the score in each test and the order of finishing. These two correlations were respectively $-.13 \pm .07$, $-.12 \pm .07$. The

similarity of the results in the two tests is striking. It appears rather clear, therefore, that so far as this situation is concerned, the speed of performance and the quality of performance are not related.

It is interesting to speculate on the possible factors which affect speed and quality of work and the relation between them. What the bearing of these factors is we cannot say with any assurance. Future investigation may isolate them and indicate their relative importance. It is also possible that still other factors may be discovered.

In the first place, it is possible that high degree of ability may operate to render high quality of performance possible and also to affect favorably the speed of performance. In so far as this is true there should be a positive correlation between speed and quality. The most common view has probably been that this hypothesis is the correct one. In the second place the adoption of a good method of procedure or training in good method should probably both raise the quality of performance and increase its speed. Poor training or lack of training would operate in the opposite direction. Such a factor as this would again produce a correlation.

While these two factors may exist, it is clear that they cannot exist alone, in the light of the fact that the correlation is sometimes not present. We may therefore set up certain other hypotheses. For example, there may be a native difference in the rate of working or in the rate of activity in general and this may be relatively independent of the quality or work of which one is capable. Observation seems to indicate that this is true, at least to some degree. There are some persons who act slowly, but who seem capable of performing difficult intellectual feats, and on the other hand there are persons who are very rapid in their responses, both mental and physical, but whose mental capacities are very limited. Similarly there may be real differences in the quality of work of which one is capable which are independent of speed of performance. This is the other side of the same situation.

In addition to these possible fundamental characteristics of intellectual makeup, there may be differences in ways of working which are more or less accidental. For example, unusual carefulness in mental work may produce a somewhat higher quality than would otherwise be produced, but may do so at the expense of speed. On the other hand a person may work recklessly and in this way speed up at the expense of quality. Such differences in method may be due not to fundamental differences in the nature, but to differences in the type of training or in temporary attitude.

The relation between speed and accuracy may further be affected by the difficulty of the test itself. It is conceivable at least that a test with a wide range of difficulty which is easy enough to permit those of little ability to make a score, but difficult enough to tax the ability of the ablest would give less correlation between speed and quality than a test which is easy and in which the score was determined chiefly by the speed with which one worked. This hypothesis suggests a beginning point for experimentation on this problem. If one should take the same group of individuals and find the correlation between speed and quality with a variety of tests, some of which

are easy and some difficult, he could determine whether or not this hypothesis had any ground.

The whole problem is one which has more than merely theoretical importance. It is important in the first place for interpretation of the results of tests. If there is low correlation between speed and quality a speed test may do injustice to certain types of pupils. If there is high correlation on the other hand, such a test would at least do much less injustice. In the second place, if there are marked differences in the type of performance with reference to speed and quality of different persons this may have an important bearing upon the kind of vocation to which various individuals are suited. Some vocations may demand rapidity to a much greater extent than others. On the other hand the quality of performance in some cases may greatly outweigh the importance of speed. Further investigation of this problem therefore promises to be not only theoretical, but of practical importance.

School Surveys

Under Direction of the Department of Measurement and Research
University of Nebraska

Doctor Fordyce has recently sent out circulars which give directions for making surveys. The following paragraphs indicate the nature of the work which is being carried on:

"A number of superintendents in Nebraska are engaged this year in making a critical study of their local school system. This study embraces the curriculum; achievements of the pupils; quality of instruction; supervision; classification and progress; social activities; school hygiene; selection, training, and tenure of teachers; judging the conditions of the buildings and equipments, through the use of standardized score cards; and finally the business administration including records, reports, and expenditures. These studies are to be put in the form of a detailed report, at the close of the year, and submitted to the Board of Education. Keen interest is manifest this fall in the application of the standardized scales for the purpose of guiding teachers and superintendents in determining the ability of the pupil and for adapting the course of study and the teaching to his needs."

Standardizing Tests in Fundamentals of Algebra and Geometry

W. C. Brenke, Professor of Mathematics

"For some years past these tests have been prepared by the Department of Mathematics of the University of Nebraska and sent out by the University Extension Division on the request of high schools. About seventy requests were received from as many different schools, and about one-third of these sent in reports of results which are tabulated below. Unmarked papers sent in by a number of schools are not included. In most cases the average, highest, middle, and lowest marks made by the class are given. The same serial number does not as a rule indicate the same school. This tabulation will give to the teachers who gave the test an approximate indication of the relative performance of their classes:

TABLE I. ALGEBRA

School	Average	High	Middle	Low
1	83	100	83	10
2	77	100	81	27
3	75	100	87	18
4	75	100	75	47
5	75	91	..	46
6	75	100	90	25
7	..	84	73	61
8	67	100	..	17
9	..	90	67	45
10	66	100	83	33
11	66	100	66	29
12	63	92	80	23
13	60	98	66	30
14	..	100	65	0
15	..	100	63	0
16	..	100	62	0
17	58	88	57	27
18	58	100	50	25
19	58	100	56	16
20	50	100	58	17
21	..	91	39	17

TABLE II. GEOMETRY

School	Average	High	Middle	Low
1	83	100	67	25
2	78	100	83	33
3	74	98	75	34
4	..	100	72	32
5	69	98	71	32
6	68	90	67	33
7	65	90	68	35
8	..	90	67	17
9	64	84	67	17
10	60	90	60	35
11	60	78	60	24
12	60	85	55	..
13	..	100	60	..
14	55	80	67	20
15	55	100	55	17
16	..	83	67	10
17	51	95	33	0
18	..	93	48	12
19	50	84	..	0
20	..	82	48	24
21	40	50	40	17
22	38	63	38	17